VORLESUNGSMITSCHRIFT

WAHRSCHEINLICHKEITSTHEORIE II
– STOCHASTISCHE PROZESSE –

WINTERSEMESTER 2013/2014,
UNIVERSITÄT LEIPZIG

MAX VON RENESSE
AUFGEZEICHNET VON ROBERT GALLE

VORLÄUFIGE VERSION

SEPTEMBER 7, 2015

Inhaltsverzeichnis

1. Langzeitverhalten von Markovketten

1.1. **Markovketten – Wiederholung grundlegender Definitionen.** Wir knüpfen an die letzten Abschnitte der Vorlesung Wahrscheinlichkeitstheorie I an, in der wir zuletzt *Markov-Ketten* $(X_k)_{k\in\mathbb{N}_0}$ auf abzählbaren Zustandsräumen $E \subset \mathbb{N}$ betrachtet haben. Hier noch einmal zur Erinnerung einige grundlegende Definitionen.

Definition 1. Es sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $(E, \mathcal{E})$ ein messbarer Raum sowie I eine Indexmenge. Dann heißt eine Familie von E-wertigen messbaren Abbildungen $(X_k)_{k\in I}$, $X_k : \Omega \to E$ ein *stochastischer Prozess.*

Bemerkung 2.

(1) Die Menge E wird dann *Zustandsraum* genannt, die Elemente von E heißen *Zustände*, messbare Funktionen $E \to \mathbb{R}$ heißen *Zustandsfunktionen* bzw. *Observablen.*

(2) Die Erwähnung eines zuvor gegebenen Wahrscheinlichkeitsmaßes P auf $(\Omega, \mathcal{F})$ ist in der Definition eigentlich überflüssig. Daher werden wir in späteren Abschnitten gelegentlich auch von stochastischen Prozessen sprechen, wenn zuvor kein (oder mehrere) Wahrscheinlichkeitsmaß(e) auf (Ω, P) fixiert wurde(n). Jede Familie von messbaren Abbildungen zwischen zwei fest gewählten messbaren Räumen kann demnach bereits als stochastischer Prozess angesehen werden.

Gelegentlich verwenden wir auch die Schreibweisen $(X_k)_{k\in I} = (X_\cdot) = X_\cdot = X$ für einen stochastischen Prozesses, sofern keine Verwechselungsgefahr besteht.

In diesem Abschnitt wollen wir uns auf den vergleichsweise einfachen Fall beschränken, wenn der Zustandsraum E *abzählbar* d.h. o.B.d.A. eine Teilmenge der natürlichen Zahlen ist.

Definition 3. Ein *Maß* auf E ist eine Funktion $\lambda : E \to \mathbb{R}_{\geq 0}$. Eine *Verteilung* auf E ist ein Maß, so dass $\sum_{i\in E} \lambda_i = 1$

Bemerkung 4. Gelegentlich werden wir von einer Verteilung auch dann sprechen, wenn nur $\sum_{i\in E} \lambda_i < \infty$ gilt. Durch *Normierung* $\bar{\lambda}_i := \frac{1}{Z}\lambda_i$ mit $Z = \sum_{i\in E} \lambda_i$ entsteht daraus dann eine Verteilung im strengen Sinne der obigen Definition.

Definition 5. Eine *stochastische Matrix* auf E ist eine Abbilung $P = (p_{ij}) : E \times E \to [0, 1]$, so dass $\sum_j p_{ij} = 1$

Eine stochastische Matrix wird auch als *Übergangskern* auf E bezeichnet, da für festes $i \in E$ die Verteilung $\lambda_j := p_{ij}, j \in E$ als Wahrscheinlichkeit für einen Sprung vom Zustand i in die Zustände $j \in E$ interpretiert wird.

Definition 6. Es sei E abzählbar, λ eine Verteilung und P eine Übergangsmatrix auf E, dann heißt ein E-wertiger stochastischer Prozess $(X_k)_{k\in\mathbb{N}_0}$ eine (λ, P)-*Markovkette*, falls

(1) $P(X_0 = i) = \lambda_i \,\forall i \in E.$

(2) $P(X_{k+1} = j | X_k = i, X_{k-1} = m_{k-1}, X_{k-2} = m_{k-2}, \cdots, X_0 = m_0) = p_{ij} \,\forall k \in \mathbb{N}_0, i, j, m_{k-1}, \ldots, m_0 \in E.$

Bemerkung 7.

(1) In diesem Fall wird λ als *Startverteilung* bezeichnet.

(2) Die Existenz einer Markov-Kette zu einem gegebenen Paar (λ, P) wurde im vergangenen Semester besprochen. Die resultierenden Wahrscheinlichkeiten bzw. Erwartungswerte für das Verhalten der Pfade der Markovkette $X.$ bei Start gemäß Startverteilung λ schreiben wir $P_\lambda(\cdots)$ bzw. $E_\lambda(\cdots)$. Der Fall eines festen (deterministischen) Startpunktes bei $X_0 = i$ für $i \in \mathbb{E}$ entspricht der Wahl $\lambda_j = \delta_i(j)$ (Dirac-Maß im Punkt $i \in E$). Wir verwenden dann die einfachere Notation $P_i(\cdots)$ und $E_i(\cdots)$.

(3) Die *elementare Markov-Eigenschaft* (2) lässt sich dann unmittelbar auf die allgemeinere Aussage erweitern, dass

$$P_\lambda(X_{k+1} \in A_1, \cdots X_{k+l} \in A_l | X_k = i, X_{k-1} \in A_{m_{k-1}}, \cdots, X_0 \in A_{m_0}) \\ = P_i(X_1 \in A_1, \cdots X_l \in A_l)$$

für $i \in E$ und beliebige Mengen $A_j \subset E$. Insbesondere ist eine Markov-Kette (X_k) *gedächtnislos* dem Sinne, dass für die Berechnung der bedingten Wahrscheinlichkeiten für künftiges Verhalten, gegeben Information über beobachtetes Verhalten in der Vergangenheit, nur die Information über den aktuellen Zustand des Prozesses (X_k) relevant ist. Die Vergangenheit beeinflußt die Wahrscheinlichkeiten für künftiges Verhalten des Prozesses also nicht, sondern nur die Gegenwart.

(4) Für die Wahrscheinlichkeiten bei n-fach wiederholtem Sprung gilt dann, dass

$$P_i(X_n = j) = P(X_n = j | X_0 = i) = (P^n)_{ij},$$

wobei $P^n = P \cdot P \cdot \cdots \cdot P$ das n-te Potenz im Sinne des Matrixproduktes $(P \cdot P)_{ij} = \sum_k p_{ik} p_{kj}$ bezeichnet. (Künftig schreiben wir $p_{ij}^{(m)}$ oder später einfach p_{ij}^m für die Einträge von P^m)

(5) Eine (λ, P)-Markovkette ist *zeitlich homogen* in dem Sinne, dass die Übergangswahrscheinlichkeiten in der elementaren Markoveigenschaft (2) oben nicht vom Zeitpunkt k abhängen. Der zeitlich inhomogenen Fall entspräche

$$P(X_{k+1} = j | X_k = i, X_{k-1} = m_{k-1}, X_{k-2} = m_{k-2}, \cdots, X_0 = m_0) = p_{ij}^{(k)}$$

für $k \in \mathbb{N}_0, i, j, m_{k-1}, \ldots, \in E$, wobei $p^{(k)}$ eine (vom Zeitpunkt k abhängige) Übergangsmatrix ist. Im folgenden wollen wir aber nur zeithomogene Ketten betrachten.

1.2. **Rekurrenz.** Im folgenden wollen wir das Langzeitverhalten von Markovketten studieren. Hierbei spielt der Begriff der Rekurrenz eine tragende Rolle.

Definition 8. Für $A \subset E$ ist die *Eintrittszeit in A* definiert als

$$\tau_+^A := \inf \{k \geq 1 \mid X_k \in A\}.$$

Bemerkung 9. Man beachte, dass der Startpunkt von $(X_k)_{k \in \mathbb{N}_0}$ bei der Definition von τ_+^A ignoriert wird. Andernfalls spricht man von einer *Trefferzeit* (vergleiche W-Theorie I). Jedoch sind unterschiedliche Bezeichnungen gebräuchlich Man sollte sich also stets vergewissern, wovon im Einzelfall die Rede ist.

Definition 10. Für $A = \{i\}$ setzen wir $\tau_i := \tau_+^{\{i\}}$. Die Folge $(s_k)_k$ der *Besuchszeiten* ist dann wie folgt definiert

$$\begin{aligned} s_0^{\{i\}} &= \tau_i \\ s_{k+1}^{\{i\}} &= \inf\left\{ l \geq s_k^{\{i\}} + 1 \mid X_l = 1 \right\}, \end{aligned}$$

mit der Konvention, dass $\inf \emptyset = \infty$.

D.h. falls $s_k^{\{i\}} < \infty$ besucht die Kette zum Zeitpunkt $s_k^{\{i\}}$ den Zustand i zum k+1-ten Mal, wobei ein möglicher Start in i nicht als Besuch gewertet wird. – Zur Einfachheit lassen wir im ~~folgenden~~ den oberen Index $\{i\}$ weg, wenn keine Gefahr einer Verwechselung besteht.

Definition 11. Für $i \in E$ heißt

$$f_i = P_i(\tau_i < \infty)$$

die *Wahrscheinlichkeit der Wiederkehr (in endlicher Zeit) des Zustands i.*

Wir wollen auch die Anzahl ~~der Besuche~~ Wiederkehren eines Zustands $i \in E$ durch die Markovkette zählen wie folgt

$$N_i = \sum_{l=1}^{\infty} 1_{\{i\}}(X_l).$$

Wir werden gleich sehen, dass bei Start der Kette $(X_n)_n$ in i die ~~Besuchszahl~~ Wiederkehrzahl N_i eine geometrisch verteilte Zufallsvariable ist mit Parameter f_i. Dafür erinnern wir noch einmal an die geometrische Verteilung.

Definition 12. Sei Z eine $\mathbb{N}_0 \cup \{\infty\}$-wertige Zufallsvariable. Sie heißt *geometrisch verteilt* mit Parameter $r \in [0,1]$, wenn

$$P(Z = k) = \begin{cases} r^k(1-r) & \text{für } k \in \mathbb{N}_0 \\ 1 - (1-r)\sum_{k\in\mathbb{N}_0} r^k & \text{für } k = \infty. \end{cases}$$ 0 bzw. 1

Bemerkung 13.

(1) Eine geometrisch verteilte Zufallsvariable entsteht z.B. als die Anzahl der Erfolge bis zum ersten Misserfolg beim wiederholten (stochastisch unabhängigen) Werfen einer Münze mit Erfolgsparameter r.

(2) Die obige Definition für $P(Z = \infty)$ bedeutet schlicht, dass je nachdem ob $r < 1$ oder $r = 1$ für eine geometrisch mit Parameter $r \in [0,1]$ verteilte Zufallsvariable Z gilt, dass $Z < \infty$ fast sicher bzw. $Z = \infty$ fast sicher. Inbsesondere sind äquivalent

$$r < 1 \Leftrightarrow Z < \infty \text{ f.s.} \Leftrightarrow \mathbb{E}(Z) < \infty.$$ $\mathbb{E}(Z) = \frac{1}{1-r}$

Angesichts der Interpretation einer geometrisch verteilten Zufallsvariable als Anzahl der Erfolge beim Münzwurf vor dem ersten Misserfolg ist das nächste Resultat sehr intuitiv.

Satz 14. *Falls die Kette $(X_n)_{n\in\mathbb{N}_0}$ im Zustand $i \in E$ startet, ist N_i geometrisch verteilt mit Parameter f_i, d.h. für $k \in \mathbb{N}_0$ gilt*

$$P_i(N_i = k) = f_i^k \cdot (1 - f_i).$$

Zum Beweis benötigen wir das folgende Lemma.

Lemma 15. *Mit den Definitionen wie oben gilt*

$$P_i(N_i \geq k+1 | s_{k-1} = n) = f_i.$$

Beweis. Die Aussage ist intutiv klar und folgt einfach zudem aus der *starken Markov-Eigenschaft*, die wir jedoch erst ~~sehr viel~~ später einführen wollen. – Wir geben daher an dieser Stelle einen ausführlichen elementaren Beweis. Nach Definition der Folge (s_k) ist die Aussage des Lemmas äquivalent zu $P_i(s_k < \infty | s_{k-1} = n) = P_i(s_0 < \infty)$, daher reicht es zu zeigen, dass für alle $l \in \mathbb{N}$

$$P_i(s_k = n + l | s_{k-1} = n) = P_i(s_0 = l).$$

Dass dies eine Konsequenz der Markov-Eigenschaft von (X_n) ist, sieht man wie folgt. Das Ereignis

$$\{s_{k-1} = n\} = \dot{\bigcup}_{1 \leq t_1 \leq t_2 \leq \cdots t_{k-1} < n} E_{t_1, \cdots, t_{k-1}}$$

lässt sich als disjunkte Vereinigung darstellen von Ereignissen der Form

$$E_{t_1, \cdots, t_{k-1}} = \left\{ X_{t_1} = X_{t_2} = \cdots = X_{t_{k-1}} = X_n = i, X_m \neq i \forall m \in \{1, \cdots, n-1\} \setminus \{t_1, \cdots t_{k-1}\} \right\}.$$

Insbesondere folgt aus dieser Zerlegung zunächst einmal, dass

$$P_i(s_k = n + l | s_{k-1} = n) = \sum_{1 \leq t_1 \leq t_2 \leq \cdots t_k < l} P_i(s_k = n + l | E_{t_1, \cdots, t_{k-1}}) \cdot P_i(E_{t_1, \cdots, t_{k-1}} | s_{k-1} = n).$$

Aufgrund der Markov-Eigenschaft von (X_n) geht nun nur die Bedingung $X_n = i$ aus der Definition von $E_{t_1, \cdots, t_{k-1}}$ in die Berechnung der bedingten Erwartung $P_i(s_k = n + l | E_{t_1, \cdots, t_{k-1}})$ ein, d.h.

$$\begin{aligned} P_i(s_k = n + l | E_{t_1, \cdots, t_{k-1}}) &= P_i(X_{n+1} \neq i, \cdots, X_{n+l-1} \neq i, X_{n+l} = i | E_{t_1, \cdots, t_{k-1}}) \\ &= P_i(X_{n+1} \neq i, \cdots, X_{n+l-1} \neq i, X_{n+l} = i | X_n = i) \\ &= P_i(X_1 \neq i, \cdots, X_{l-1} \neq i, X_l = i) \\ &= P_i(s_0 = l), \end{aligned}$$

was oben eingesetzt ergibt, dass

$$\begin{aligned} P_i(s_k = n + l | s_{k-1} = n) &= \sum_{1 \leq t_1 \leq t_2 \leq \cdots t_k < l} P_i(s_0 = l) \cdot P_i(E_{t_1, \cdots, t_{k-1}} | s_{k-1} = n) \\ &= P_i(s_0 = l). \end{aligned}$$

□

Beweis von Satz 14. Die Aussage ist äquivalent zu $P_i(N_i \geq k) = f_i^k$ für alle $k \in \mathbb{N}_0$. Wir beweisen letzteres durch Beweis durch Induktion nach k. Für $k = 0$ ist die Aussage offensichtlich richtig. Für $k+1$ gilt

$$\begin{aligned} P_i(N_i \geq k + 1) &= P_i(N_i \geq k + 1, N_i \geq k) \\ &= P_i(N_i \geq k + 1, s_{k-1} < \infty) \\ &= P_i(N_i \geq k + 1 | s_{k-1} < \infty) P_i(s_{k-1} < \infty) \\ &= P_i(N_i \geq k + 1 | s_{k-1} < \infty) P_i(N_i \geq k) \\ &= P_i(N_i \geq k + 1 | s_{k-1} < \infty) f_i^k, \end{aligned}$$

wobei wir im letzten Schritt die Induktionsvoraussetzung benutzt haben. – Für die Bestimmung der obenstehenden bedingten Wahrscheinlichkeit folgern

Bemerke $P(A \mid \dot{\bigcup} E_i) = \sum P(A \mid E_i)\, P(E_i \mid \dot{\bigcup} E_i)$

wir mit der Darstellung des Ereignisses $\{s_{k-1} < \infty\} = \dot{\bigcup}_{n\in\mathbb{N}}\{s_{k-1} = n\}$ als disjunkte Vereinigung schließlich

$$\begin{aligned} P_i(N_i \geq k+1 | s_{k-1} < \infty) &= \sum_{n\in\mathbb{N}} P_i(N_i \geq k+1 | s_{k-1} = n) P_i(s_{k-1} = n | s_{k-1} < \infty) \\ &= \sum_{n\in\mathbb{N}} f_i P_i(s_{k-1} = n | s_{k-1} < \infty) \\ &= f_i, \end{aligned}$$

und erhalten somit die Behauptung. □

Korollar 16. *Für einen Zustand $i \in E$ gilt*

$$N_i < \infty\, P_i\text{-fast sicher} \Leftrightarrow E_i(N_i) < \infty \Leftrightarrow f_i < 1 \Leftrightarrow \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty.$$

Beweis. Die ersten zwei Äquivalenzen folgen aus der Tatsache, dass $Z := N_i$ bei Start in i geometrisch verteilt ist. Die letzte Äquivalenz folgt aus

$$\mathbb{E}_i\left(N_i\right) = \mathbb{E}\left(\sum_{l=1}^{\infty} 1_i\left(X_l\right)\right) = \sum_{l=1}^{\infty} P_i\left(X_l = i\right) = \sum_{l=1}^{\infty} p_{ii}^{(l)}.$$

□

Definition 17. Ein Zustand $i \in E$ heißt

$$\begin{aligned} \textit{transient} &\iff P_i(X_n = i \text{ für unendlich viele n}) = 0 \\ \textit{rekurrent} &\iff P_i(X_n = i \text{ für unendlich viele n}) = 1 \end{aligned}$$

Gemäß Korollar 16 finden wir die Dichotomie, dass entweder $N_i < \infty$ oder $N_i = \infty$ fast sicher unter P_i, was die nächste Definition rechtfertigt. Insbesondere gilt der folgende Satz.

Korollar 18. *Ein Zustand ist rekurrent genau dann, wenn $P_i(\tau_i < \infty) = 1$, andernfalls ist er transient.*

Rekurrenz von Zuständen ist eine Klasseneigenschaft. Hierzu erinnern wir an den Begriff der (kommunizierenden) Klassen.

Definition 19. Für $i, j \in E$ heißt j erreichbar von i, falls $p_{ij}^m > 0$ für ein geeignete gewähltes $m \in \mathbb{N}$. Die von der Äquivalenzrelation

$$i \sim j :\Leftrightarrow \text{i erreichbar von } j \text{ und } j \text{ erreichbar von } i$$

induzierten Äquivalenklassen heißen *kommunizierende Klassen.*

transient

Korollar 20. *Für zwei Zustände $i \sim j$ derselben Klasse ist i ~~rekurrent~~ genau dann, wenn j ~~rekurrent~~ ist.*

transient

Beweis. Dies folgt aus $p_{jj}^n \geq p_{ij}^m p_{ii}^{n-m-m'} p_{ji}^{m'}$ für $n \geq m + m'$, so dass $\sum_n p_{jj}^n < \infty \Rightarrow \sum_m p_{ii}^m < \infty$. □

Insbesondere ist die Rekurrenz im irreduziblen Fall eine globale Eigenschaft der Markov-Kette. (Eine Markovkette heißt *irreduzibel*, falls E aus genau einer kommunizierenden Klasse besteht, d.h. wenn jeder Zustand von jedem anderen Zustand erreichbar ist.).

Beispiel 21 (Einfache Irrfahrt auf $\mathbb{Z}^d$)**.** Die einfache Irrfahrt auf $\mathbb{Z}^d$ ist die Markovkette auf $E := \mathbb{Z}^d$ mit $p_{ij} = \frac{1}{2^d}$, falls i, j benachbart in $\mathbb{Z}^d$ sind, bzw. $p_{ij} = 0$ sonst. Dieses Beispiel hatten wir im letzten Semester diskutiert und festgestellt, dass für große n gilt

$$p_{ii}^n \begin{cases} \sim n^{-\frac{d}{2}} \text{ falls } n \text{ gerade,} \\ = 0 \text{ falls } n \text{ ungerade.} \end{cases}$$

Somit ist die einfache Irrfahrt aus $\mathbb{Z}^d$ rekurrent für $d = 1, 2$ und transient für $d \geq 3$.

1.2.1. *Invariante Maße und Rekurrenz.*

Definition 22. $\lambda : E \to \mathbb{R}_+$ heißt *invariantes Maß*, falls

$$\lambda_i = \sum_{j \in E} \lambda_j p_{ji} \quad \forall i \in E.$$

gilt. Ein invariantes Maß λ heißt *invariante Verteilung*, falls zudem gilt, dass $\sum_{i \in E} \lambda_i < \infty$.

Bemerkung 23.

(1) Jede Übergangsmatrix P auf dem Zustandsraum E definert eine Operation auf den Maßen auf E wie folgt

$$\nu \mapsto \nu P \text{ mit } (\nu P)_i := \sum_{j \in E} \nu_j p_{ji}.$$

Falls ν ein Wahrscheinhlichkeitsmaß auf E ist, so lässt sich die Zahl $(\nu P)_j = \sum_{j \in E} \nu_j p_{ji}$ interpretieren als Wahrscheinlichkeit, bei einem zweifachen Sprung (zunächst in einen gemäß ν zufällig gewählten Anfangszustand E hinein und von dort dann in einem zweiten Schritt als einfachem Sprung gemäß dem Übergangskern P) im Zustand i zu landen. νP ist also die Verteilung, die nach einem einmaligen 'Durchmischen' der Ausgangsverteilung ν gemäß P entsteht. Die P-Invarianz von λ besagt dann, dass λ sich hierbei nicht ändert.

(2) Falls λ eine Verteilung ist, so auch

$$\overline{\lambda}_i := \frac{\lambda(i)}{\sum_{j \in E} \lambda(j)}.$$

Dann ist $\overline{\lambda} : E \to \mathbb{R}_+$ ebenfalls invariant mit

$$\sum_{i \in E} \overline{\lambda}(i) = 1$$

(3) In der Sprache der linearen Algebra ist ein invariantes Maß also ein Links-Eigenvektor $(\lambda_i)_{i \in E} \in \mathbb{R}^E$ von P mit nichtnegativen Einträgen und Eigenwert 1. Insbesondere ist die Konvexkombination von invarianten Maßen wieder ein invariantes Maß.

Die Existenz von nicht-trivialen invarianten Maßen ist keine Selbstverständlichkeit. Falls wir wenigstens einen rekurrenten Zustand finden, können wir hiermit mindestens ein invariantes Maß konstruieren wie folgt.

Satz 24. *Falls $k \in E$ rekurrent, dann ist ein invariates Maß gegeben durch*

$$E \ni i \to \gamma_i^{(k)}, \quad \gamma_i^{(k)} := \mathbb{E}_k\big(\sum_{l=1}^{\tau_k} 1_{\{i\}}(X_l)\big)$$

$\gamma_i^{(k)}$ = Exp. Anz. Besuche von i vor Wiederkehr nach k.

Bemerkung 25. Für einen beliebigen Zustand $k \in E$ gilt

$$\gamma_k^{(k)} = \mathbb{E}\left(1_{\tau_k < \infty}\right) = P_k(\tau_k < \infty),$$

d.h. k ist Rekurrent genau dann, wenn $\gamma_k^{(k)} = 1$.

Beweis von Satz 24. Für $i \in \mathbb{E}$ beliebig ist

$$\begin{aligned}
\gamma_i^{(k)} &= \mathbb{E}_k(\sum_{l=1}^{\tau_k} 1_{\{i\}}(X_l)) \\
&= \mathbb{E}_k(\sum_{l=1}^{\infty} 1_{\{i\}}(X_l) 1_{\{l \le \tau_k\}}) \\
&= \sum_{l=1}^{\infty} \mathbb{E}_k(1_{\{i\}}(X_l) 1_{\{l \le \tau_k\}}) \\
&= \sum_{l=1}^{\infty} P_k(X_l = i; l \le \tau_k) \\
&= \sum_{l=2}^{\infty} P_k(X_l = i; l \le \tau_k) + P_k(X_1 = i; \tau_k \ge 1).
\end{aligned}$$

Semikolons durch Kommatas ersetzen

Für den zweite Summanden gilt wegen $\tau_k \ge 1$ und der Rekurrenz von k, dass

$$P_k(X_1 = i; \tau_k \ge 1) = P_k(X_1 = i) = p_{ki} = p_{ki}\gamma_k^{(k)}.$$

Für den ersten Summanden schreiben wir

$$
\begin{aligned}
&\sum_{l=2}^{\infty} P_k(X_l = i; l \leq \tau_k) \\
&\quad = \sum_{l=2}^{\infty} \sum_{j \in E} P_k(X_l = i; X_{l-1} = j; l \leq \tau_k) \\
&\quad = \sum_{l=2}^{\infty} \sum_{j \in E \setminus \{k\}} P_k(X_l = i; X_{l-1} = j; l \leq \tau_k) \\
&\quad = \sum_{l=2}^{\infty} \sum_{j \in E \setminus \{k\}} P_k(X_1 \neq k, ..., X_{l-1} \neq k, X_l = i, X_{l-1} = j) \\
&\quad = \sum_{l=2}^{\infty} \sum_{j \in E \setminus \{k\}} P_k(X_l = i \mid X_1 \neq k, ..., X_{l-1} \neq k, X_{l-1} = j) \\
&\quad \times P_k(X_1 \neq k, ..., X_{l-1} \neq k, X_{l-1} = j) \\
&\quad = \sum_{l=2}^{\infty} \sum_{j \in E \setminus \{k\}} \underbrace{P(X_l = i \mid X_{l-1} = j)}_{= p_{ji}} P_k(X_1 \neq k, ..., X_{l-1} \neq k, X_{l-1} = j) \\
&\quad = \sum_{l=2}^{\infty} \sum_{j \in E \setminus \{k\}} p_{ji} P_k(X_{l-1} = j, \tau_k \geq l-1) \\
&\quad = \sum_{j \in E \setminus \{k\}} p_{ji} \sum_{l=2}^{\infty} P_k(X_{l-1} = j; \tau_k \geq l-1) \\
&\quad = \sum_{j \in E \setminus \{k\}} p_{ji} \sum_{l=1}^{\infty} P_k(X_l = j; \tau_k \geq l) \\
&\quad = \sum_{j \in E \setminus \{k\}} P_{ji} \gamma_j^{(k)}
\end{aligned}
$$

Fasst man beide Summen zusammen, so ergibt sich hieraus $(\gamma^{(k)} P)_i = \gamma_i^{(k)}$. □

Beispiel 26. (1) Seien a, b, c drei Zustände mit

$$
\begin{array}{ccc}
a & \xrightarrow{p=1} & b \\
b & \xrightarrow{1} & a \\
c & \xrightarrow{\frac{1}{2}, \frac{1}{2}} & a, b
\end{array}
$$

Diese Kette ist nicht irreduzibel. a, b sind rekurrent, c ist transient und

$$
\begin{aligned}
\gamma_c^{(a)} &= 0 \\
\gamma_{a,b}^{(a)} &= 1
\end{aligned}
$$

(2) Seien nun a, b, c, d, e gegeben mit

$$\begin{array}{ccc} a & \overset{1}{\longleftrightarrow} & b \\ c & \overset{1}{\longleftrightarrow} & d \\ e & \overset{\frac{1}{2},\frac{1}{2}}{\longrightarrow} & b, d \end{array}$$

Die ersten zwei Beziehungen bilden kommunizierende Klassen. Wähle beispielsweise $k = a$. Dann:

$$\gamma^{(k)} = (1, 1, 0, 0, 0)^T$$

und falls $k = c$:

$$\gamma^{(k)} = (0, 0, 1, 1, 0)^T$$

Die γ sind invariante Maße. Ebenfalls invariant ist dann:

$$\alpha(1, 1, 0, 0, 0)^T + (1 - \alpha) \cdot (0, 0, 1, 1, 0)^T \forall \alpha \in [0, 1]$$

Das folgende Lemma zeigt im irreduziblen Fall, dass das in Satz 24 konstruierte invariante Maß nicht-trivial (d.h. ungleich dem Null-Maß) und zudem nicht-degeneriert (d.h. nirgendwo unendlich) ist.

Lemma 27. *Falls k rekurrent und P irreduzibel sind, dann ist $\gamma_i^{(k)} \in]0, \infty[$.*

Beweis. $\forall i \in E \exists m : p_{ki}^{(m)} > 0$.

$$\implies \gamma_i^{(k)} = \sum_{j \in E} P_{ji}^{(m)} \gamma_j \geq P_{ki}^{(m)} \cdot 1 > 0.$$

Andererseits ~~Analog~~

$$\gamma_k^{(k)} = 1 = \sum_{j \in E} p_{jk}^{(n)} \gamma_j^{(k)} \geq p_{ik}^{(n)} \gamma_i^{(k)}$$

$$\implies \gamma_i^{(k)} \leq \frac{1}{p_{ik}^{(n)}} < \infty.$$

□

Aussage und Beweis des vorigen Lemmas erlauben unmittelbar die folgende Verallgemeinerung.

Satz 28. *Es sei P eine irreduzible Übergangsmatrix E und λ ein invariantes Maß. Dann ist λ entweder trivial ($\lambda_i = 0 \forall i \in E$) oder strikt positiv ($\lambda_i \in]0, \infty[\forall i \in \mathbb{E}.$).*

$E = \mathbb{N}$

Beispiel 29 (Fluchtkette)**.** Es seien $P_i, i \leq 0$ die Wahrscheinlichkeiten entlang der reellen Achse zu springen, also von Null aus ist P_0 die WKeit nach rechts zu springen, dann P_1 um wieder nach rechts. Die GegenWKeiten entsprechend $1 - P_i$. Wir nehmen an, dass

$$P := \prod_{l=0}^{\infty} P_l > 0.$$

Wir ermitteln die

Invariante Verteilung:

$$\pi_j = \pi_{j-1} P_j \implies \pi_i = \left(\prod_{l=0}^{i-1} P_l \right) \pi_0$$

An der Stelle Null folgt:

$$\begin{aligned}
\pi_0 &= \left(\sum_{i\in\mathbb{N}_0}\left(\prod_{l=0}^{i-1}P_l\right)(1-p_i)\right)\pi_0 \\
&= \pi_0\left(\sum_{i\in\mathbb{N}_0}\left(\prod_{l=0}^{i}P_l-\prod_{l=0}^{i-1}P_l\right)\right) \\
&= \pi_0(P-1) \\
\Longrightarrow \pi_0 &= 0 \\
\Longrightarrow \pi &\equiv 0.
\end{aligned}$$

Somit existiert kein nicht-triviales invariantes Maß, die Kette kann also nicht rekurrent sein.

1.2.2. *Irreduzibilität und Eindeutigkeit des Invarianten Maßes.* Für irreduzible Ketten ist das invariante Maß bis auf multiplikative Konstante eindeutig bestimmt, wie wir nun sehen werden.

Satz 30. *Falls ~~P irreduzibel und~~ λ ein invariantes Maß mit $\lambda_k = 1$ für ein $k \in E$ sind, so gilt $\gamma_i^{(k)} \leq \lambda_i$ für alle $i \in E$.*

Beweis. Für $j \neq k$ gilt

$$\begin{aligned}
\lambda_j &= \sum_{l\in E}\lambda_l p_{lj} \\
&= \lambda_k p_{kj} + \sum_{l_1\in E\setminus\{k\}} p_{l_1 j}\lambda_{l_1} \\
&= p_{kj} + \sum_{l_1\in E\setminus\{k\}}\sum_{l_2\in E} p_{l_1 j}p_{l_2 l_1}\lambda_{l_2} \\
&= p_{kj} + \sum_{l_1\in E\setminus\{k\}} p_{l_1 j}p_{k l_1} + \sum_{l_1\in E\setminus\{k\}}\sum_{l_2\in E\setminus\{k\}} p_{l_1 j}p_{l_2 l_1}\lambda_{l_2} \\
&\vdots \\
&= p_{kj} + \sum_{l_1\in E\setminus\{k\}} p_{k l_1}p_{l_1 j} + \sum_{\substack{l_1\in E\setminus\{k\}\\ l_2\in E\setminus\{k\}}} p_{j l_1}p_{l_2 l_1}p_{k l_2} + \ldots + \sum_{\substack{l_n\in E\\ l_i\neq k, i=1,\ldots,n-1}} p_{l_1 j}p_{l_2 l_1}\cdots p_{l_n l_{n-1}}\lambda_{l_n} \\
&\geq p_{kj} + \sum_{l_1\in E\setminus\{k\}} p_{k l_1}p_{l_1 j} + \sum_{\substack{l_1\in E\setminus\{k\}\\ l_2\in E\setminus\{k\}}} p_{j l_1}p_{l_2 l_1}p_{k l_2} + \ldots + \sum_{l_i\neq k, i=1,\ldots,n-1} p_{l_1 j}p_{l_2 l_1}\cdots p_{k l_{n-1}} \\
&= \sum_{r=1}^{n} P_k\left(X_r = j; X_1\neq k, .., X_{r-1}\neq k\right) = \mathbb{E}\left(\sum_{r=1}^{n} 1_{\{j\}}(X_r)\,1_{\{\tau_k\leq r\}}\right) \\
&\overset{n\to\infty}{\longrightarrow} \mathbb{E}_k\left(\sum_{r=1}^{\infty} 1_{\{j\}}\left(X_r\right)1_{\{\tau_k\leq r\}}\right) = \gamma_j^{(k)}
\end{aligned}$$

$\square$

Korollar 31. *Falls P irreduzibel und rekurrent ist, dann ist $(\gamma_i^{(k)})_{i\in E}$ das bis auf Multiplikation mit einer Konstanten eindeutig bestimmte nichttriviale invariante Maß.*

Beweis. Es sei λ ein nichttriviales invariante Maß, dann existiert insbesondere ein $k \in E$ mit $\lambda_k > 0$. Durch Multiplikation mit einer Konstanten können wir o.b.d.A. davon ausgehen, dass $\lambda_k = 1$. Das vorausgegangene Lemma liefert die Aussage, dass $\gamma_.^{(k)} \leq \lambda_.$ auf E. Daher definiert $\nu_. := \lambda_. - \gamma_.^{(k)}$ ein weiters invariantes Maß auf E. Da $\nu_k = 0$ folgt mit Satz 28, dass $\nu_i = 0$ für alle $i \in E$. $\square$

??

Beispiel 32. Die assymmetrische Irrfahrt auf $\mathbb{Z}$ ist irreduzibel. Das invariante Maß läßt sich explizit bestimmen wie folgt.

1.2.3. *Invariante Verteilungen und Positive Rekurrenz.* Die folgende Eigenschaft stellt eine Verschärfung der Rekurrenz eines Zustands dar.

Definition 33. Ein Zustand $k \in E$ heißt *positiv rekurrent*, falls

$$\mathbb{E}_k(\tau_k) < \infty.$$

Bemerkung 34. Ein postiv rekurrenter Zustand ist insbesondere rekurrent, denn eine integrierbare nichtnegative Zufallsvariable ist notwendig endlich fast sicher. Die Umkehrung ist i.A. falsch, denn eine fast sicher endliche Zufallsvariable ist nicht notwendig integrierbar (wie das einfache Beispiel der Zufallsvariablen $X : [0,1] \mapsto \mathbb{R} \cup \{\infty\}$, $\omega \mapsto X(\omega) := \frac{1}{\omega}$ auf dem Einheitsinvervall zeigt).

Bemerkung 35. Ein rekurenter aber nicht positiv rekurenter Zustant $k \in E$ heißt auch *null-rekurrent*.

Der folgende Satz für positiv rekurrente Ketten liefert eine besonders anschauliche Darstellung des invarianten Maßes eines Zustands als Kehrwert der mittleren Dauer der Wiederkehr.

Satz 36. *Sei P irreduzibel auf E, dann sind äquivalent:*

1. *k positiv rekurrent für alle $k \in E$*
2. *$\exists k \in E$ mit k positiv rekurrent*
3. *$\exists$ invariante Verteilung auf E, d.h. π invariant und $\sum_{i\in E} \pi_i < \infty$.*

In dem Fall ist $\pi_i = \frac{1}{\mathbb{E}_i(\tau_i)}$ ~~bis auf Multiplikation~~ eindeutig.

Beweis. 2)$\Rightarrow$3): Sei k positiv rekurrent. Dann:

$$\begin{aligned}
\sum_{i\in E} \gamma_i^{(k)} &= \sum_{i\in E} \mathbb{E}_k\left(\sum_{l=1}^{\tau_k} 1_{\{i\}}(X_l)\right) \\
&= \mathbb{E}\left(\sum_{l=1}^{\tau_k} \underbrace{\sum_{i\in E} 1_{\{i\}}(X_l)}_{=1}\right) \\
&= \mathbb{E}_k(\tau_k) < \infty
\end{aligned}$$

Dann ist $\pi_i := \pi_i^{(k)}$ ein invariantes Maß mit summierbaren Einträgen, also ist π_i eine invariante Verteilung.

3)⇒1): Sei $\pi = (\pi_i)_{i\in E}$ eine nichttriviale invariante Verteilung. Dann ist nach Satz 28 $\pi_j > 0$ für alle $j \in E$. Sei $k \in E$. Dann gilt für

$$\lambda_i := \frac{\pi_i}{\pi_k}$$

nach Satz 30, dass

$$\lambda_i \geq \gamma_i^{(k)} \forall i \in E.$$

Somit

$$\mathbb{E}_k\left(\tau_k\right) = \sum_{i\in E} \gamma_i^{(k)} \leq \sum_{i\in E} \lambda_i = \frac{1}{\pi_k} \sum_{i\in E} \pi_i < \infty,$$

d.h. der Zustand k ist positiv rekurrent. Mithin ist die gesamte Kette rekurrent, und somit gemäß Korollar 31 das invariante Maß π bis auf Multiplikation mit einer Konstanten eindeutig. Wegen $\lambda_k = \gamma_k^{(k)}$ ist also $\gamma_\cdot^{(k)} = \lambda_\cdot$, so dass die Ungleichung oben sogar eine Gleichung ist, d.h.

$$\mathbb{E}_k\left(\tau_k\right) = \frac{1}{\pi_k} \underbrace{\sum_{i\in E} \pi_i}_{=C} = 1$$

$$\Longrightarrow \pi_k := \frac{C}{\mathbb{E}_k\left(\tau_k\right)}$$

□

Beispiel 37. Wir haben gesehen, dass die symmetrische Irrfahrt auf $\mathbb{Z}$ rekurrent ist. Wegen der Irreduzibilität ist das invariante Maß bis auf multiplikative Konstante eindeutig bestimmt. Das konstante Maß auf $\mathbb{Z}$, d.h. $\lambda_i = 1$ für alle i ist invarant für die symmetrische Irrfahrt, aber nicht endlich, denn $\sum_{i\in\mathbb{Z}} \lambda_i = \sum_{i\in\mathbb{Z}} 1 = \infty$. Daher ist die symmetrische Irrfahrt auf $\mathbb{Z}$ zwar rekurrent aber nicht positiv rekurent.

1.3. **Ergodensatz für Markovketten.** Dass die invariante Verteilung in der Tat Information über das Langzeitverhalten der Trajektorien enthält, liefert der folgende Satz. Er besagt, dass das entlang von Trajektorien pfadweise gebildete Zeitmittel von Observablen für lange Beobachtungszeiträume gegen den gemäß der invarianten Verteilung gebildeteten Mittelwerte der Observablen (im Sinne der fast sicheren Konvergenz von Zufallsvariablen) konvergiert. Hierbei nenen wir eine Abbbildung $f : E \to \mathbb{R}$ *Observable* oder *Zustandsfunktion*.

Satz 38 (Erdogensatz für Markovketten)**.**

(1) *Sei* (X_k) *eine irreduzible Markovkette auf* E*, dann gilt:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} 1_i\left(X_l\right) = \pi_i := \frac{1}{\mathbb{E}_i(\tau_i)} \text{ fast sicher.}$$

(2) *Falls* (X_k) *zudem positiv rekurrent ist, so gilt für alle* $f : E \to \mathbb{R}$ *beschränkt*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} f\left(X_l\right) = \sum_{k\in E} f(k)\pi_k \text{ fast sicher.}$$

Bemerkung 39. (1) Der Satz kann als eine Variante des starken Gesetztes der großen Zahlen aufgefasst werden für eine Folge von (nicht unabhängigen) Zufallsvariablen, die als Observable entlang der realisierten Trajektorien einer Markovkette zustande kommen.
(2) Der Grenzwert ist stets derselbe, unabhängig von der Startverteilung von $(X_k)_k$.

Für den Beweis von Satz 38 benutzen wir das starke Gesetz der Großen Zahlen für eine geeignete Folge von Zufallsvariablen. Zu diesem Zweck erinnern wir an in Definition 10 eingeführte Folge von Besuchszeiten $(s_n^{(i)})_{n\in\mathbb{N}_0}$ des Zustands i. In Erweiterung hiervon sei

$$\tau_0^{(i)} := \tau_i, \quad \tau_n^{(i)} := s_n^{(i)} - s_{n-1}^{(i)}, \quad n \in \mathbb{N},$$

die Folge der *Wartezeiten zwischen den aufeinanderfolgenden Besuchen* des Zustands i.

Lemma 40. *Die Folge $(\tau_n^{(i)})_{n\in\mathbb{N}}$ ist unabhängig identisch verteilt mit $E(\tau_n^{(i)}) = E_i(\tau_i)$*

Beweis. Die Aussage ist eine Konsequenz der Markov-Eigenschaft von $(X_k)_{k\in\mathbb{N}_0}$ und lässt sich ähnlich zum zum Beweis von Lemma 15 elementar beweisen. Man beachte hierzu, dass die Startverteilung von $(X_k)_{k\in\mathbb{N}_0}$ allein in die Verteilung der Zufalsvariable $\tau_0^{(i)}$ eingeht. Von Letzterer ist im Lemma jedoch keine Rede. (Details siehe Übung, siehe auch die folgende Bemerkung 41.) □

Bemerkung 41. Die Pfade der Markovkette zwischen zwei Besuchen eines Zustands i werden auch *Exkursionen* genannt. Die Folge der Exkursionen sind ebenfalls stochastisch unabhängig. – Falls i rekurrent ist, kann die gesamte Markovkette durch sukzessiven Ausführung von stochastisch unabhängigen Exkursionen generiert werden.

Beweis von Satz 38. (1) Falls $i \in E$ transient ist, gilt $X_l = i$ für nur endlich viele l für P-fast alle $\omega \in \Omega$.

$$\Longrightarrow \frac{1}{n}\sum_{l=1}^{n} 1_{\{i\}}(X_l) \overset{n\to\infty}{\longrightarrow} 0 \text{ fast sicher}$$

Zudem ist $\mathbb{E}_i(\tau_i) = \infty$, so dass

$$\Longrightarrow 0 = \frac{1}{\infty} \iff \frac{1}{n}\sum_{i=1}^{n} 1_{\{k\}}(X_l) \overset{n\to\infty}{\longrightarrow} \frac{1}{\mathbb{E}_k(\tau_k)} = \pi_k.$$

Falls $i \in E$ rekurrent ist, schreiben wir

$$\frac{1}{n}\underbrace{\sum_{l=1}^{n} 1_{\{i\}}(X_l)}_{=V_n} = \frac{1}{n}\sum_{l=\tau_k}^{n} 1_{\{i\}}(X_l) = \frac{1}{n}V_n,$$

wobei V_n die Anzahl der Besuche des Zustandes i nach n Zeitschritten, bezeichnet.
$s_{l-1} := s_{l-1}^{(i)}$ ist der Zeitpunkt des l-ten Besuches des Zustandes i, daher gilt

$$s_{V_n-1} \le n \text{ und } s_{V_n} \ge n+1 \quad \forall n \in \mathbb{N},$$

so dass

$$\frac{n}{V_n} \geq \frac{s_{V_n-1}}{V_n} \text{ und } \frac{n+1}{V_n} \leq \frac{s_{V_n}}{V_n},$$

also schließlich

$$(1) \qquad \frac{V_n}{V_n-1}\frac{s_{V_n-1}}{V_n-1} \leq \frac{n}{V_n} \leq \frac{n\cancel{+1}}{n}\frac{s_{V_n}}{V_n}.$$

Nach dem starken Gesetz der großen Zahlen[1]

$$\frac{1}{k_n}\sum_{l=1}^{k_n}\tau_l^{(i)}(\omega) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left(\tau_1^i\right) \text{ für } P\text{-fast alle } \omega,$$

sofern $(k_n)_{n\in\mathbb{N}}$ eine Folge ist mit $k_n \to \infty$ für $n \to \infty$. Insbesondere bleibt die Aussage richtig, wenn $k_n = k_n(\omega)$ eine vom Zufallsparameter ω abhängige Folge ist. Wählen wir $k_n(\omega) = V_n(\omega)$ so gilt aufgrund der Rekurrenz von i, dass $k_n(\omega) \to \infty$ fast sicher, und daher

$$\frac{s_{V_n}}{V_n} = \frac{1}{k_n}\sum l = 1^{k_n}\tau_l^{(i)} \to \mathbb{E}(\tau_1^{(i)}) = \mathbb{E}_i(\tau_i) \text{ fast sicher,}$$

und somit wegen (1) auch

$$\frac{V_n}{n} \to \frac{1}{\mathbb{E}_i(\tau_i)} \text{ fast sicher.}$$

(2) Die Aussage folgt aus (1) für Funktionen mit endlichem Träger, d.h. vom Typ $f = \sum_{i=1}^N \alpha_i 1_{\{z_i\}}$, ~~d.h. für Funktionen~~. – Für den allgemeinen Fall von beschränktem $f : E \to \mathbb{R}$ führen wir die Folge von *empirischen Aufenthaltsmaßen* der Kette (X_k) ein

$$\pi_n(j) = \frac{1}{n}\sum_{l=1}^n 1_{\{j\}}\left(X_l\right),$$

dann gilt

$$\pi_n(j) \geq 0, \quad \sum_{j\in E}\pi_n(j) = 1,$$

d.h. π_n ist ein (zufälliges) Wahrscheinlicheitsmaß auf E. Die Aussage (2) ist dann äquivalent zu

$$\langle f, \pi_n\rangle \overset{n\to\infty}{\longrightarrow} \langle f, \pi\rangle \text{ fast sicher,}$$

wobei wir die Schreibweise $\langle f, \lambda\rangle = \sum_{i\in E} f(i)\lambda(i)$ für das Integral einer Funktion auf E gegen das Maß λ verwenden, und π die invariante Verteilung von (X_k) bezeichnet.
Es sei nun $A \subset E$ mit $\#A < \infty$, dann folgt die Aussage für $\hat{f} := f1_A = \sum_{a\in A} f(a)1_{\{a\}}$.
Aufgrund der Stetigkeit des Wahrscheinlichkeitsmaßes π auf E existiert zu $\varepsilon > 0$ $\exists A \subset E$ endlich, so dass

$$\pi(A) \geq 1 - \varepsilon.$$

[1] Das starke Gesetz der Großen Zahlen gilt für nichtnegative Zufallsvariablen auch im Fall eines unendlichen Erwartungswertes (hier also, falls $\mathbb{E}_i(\tau_i) = \infty$ gilt). In diesem Fall liefert das die Aussage, dass dann $\frac{1}{n}\sum_{l=1}^n \tau_l^{(i)} \to \infty$ fast sicher (siehe Übungen.)

Mit $\hat{f} := f 1_A$ gilt dann

$$\langle \hat{f}, \pi_n \rangle \xrightarrow{f.s.} \langle \hat{f}, \pi \rangle.$$

Mit der Dreicksungleichung erhalten wir

$$\begin{aligned} |\langle f, \pi_n \rangle - \langle f, \pi \rangle| &= \left| \underbrace{\langle f - \hat{f}, \pi_n \rangle}_{=a} + \underbrace{\langle \hat{f}, \pi_n \rangle}_{=b} - \underbrace{\langle f - \hat{f}, \pi \rangle}_{=c} - \underbrace{\langle \hat{f}, \pi \rangle}_{=d} \right| \\ &\leq |a| + |b-d| + |c| \end{aligned}$$

Für hinreichen großes n ist $|b-d| \leq \varepsilon$ beliebig klein. Falls $M = \sup_{x \in E} |f(x)|$ eine obere Schranke von f bezeichnet, gilt ferner wegen

$$f - \hat{f} = 1_{A^c} f \Longrightarrow \left| f - \hat{f} \right| \leq 1_{A^c} ||f||_\infty \leq 1_{A^c} M$$

$$\Longrightarrow |c| \leq M \langle 1_{A^c}, \pi \rangle = M \pi \left(A^c \right) \leq M \varepsilon$$

Analog:

$$|a| \leq M \langle 1_{A^c}, \pi_n \rangle = M \pi_n \left(A^c \right) = M \left(1 - \underbrace{\pi_n(A)}_{\to 1 - \varepsilon} \right) \to M \varepsilon$$

Hieraus ergibt sich schließlich

$$\limsup_{n \to \infty} |\langle f, \pi_n \rangle - \langle f, \pi \rangle| \leq M\varepsilon + 0 + M\varepsilon = 2M\varepsilon.$$

und damit die Behauptung für $\epsilon \to 0$. □

1.4. **Mischende Markovketten.** Der Ergodensatz erlaubt die Approximation der invarianten Verteilung durch Bestimmung der mittleren Aufenthaltsdauer gemäß

$$\frac{1}{n} \sum_{l=1}^{n} P\left(X_l = j \mid X_0 = i \right) = \frac{1}{n} \sum_{l=1}^{n} p_{ij}^{(n)} \longrightarrow \pi(j)$$

Man kann nun fragen, ob die zeitliche Mittelung wirklich nötig ist. Für die Berechnung durch Simulation auf einem Rechner wäre etwa etwa die stärkere Eigenschaft wünschenswert, dass bereits

$$p_{ij}^{(n)} \xrightarrow{n \to \infty} \pi(j) \quad \text{für alle } i, j \in E. \tag{2}$$

Definition 42. Eine Markovkette (X_k) heißt *mischend*, wenn (2) gilt für eine gewisse Verteilung π auf E.

Bemerkung 43. (1) Streng genommen handelt es sich hierbei um eine Eigenschaft der Übergangsmatrix P.
(2) Die Verteilung π ist dann eine invariante Verteilung für P.

Beispiel 44 (Periodische Kette in zwei Zuständen)**.** Gegeben seien zwei Zustände a und b, mit Sprüngen von a nach b und Umgekehrt jeweils mit Wahrsheinlichkeit Eins. Die resultierende Kette ist positiv rekurrent und irreduzibel, also ergodisch, d.h.

$$\pi(j) = \frac{1}{\mathbb{E}_j(\tau_j)} = \frac{1}{2}$$

Die Kette ist jedoch nicht mischend, denn

$$p_{11}^{(n)} = \begin{cases} 1 & \text{falls } n \text{ gerade} \\ 0 & \text{sonst.} \end{cases}$$

D.h. die Folge $n \to p_{11}^{(n)}$ ist nicht einmal konvergent.

Definition 45. $k \in E$ heißt

$$d_z := ggT \left\{ m \in \mathbb{N}_0 \mid p_{kk}^{(m)} > 0 \right\}$$

Periode vom Zustand k.

Bemerkung 46. Periode ist eine Klasseneigenschaft.

Definition 47. Eine irreduzible Markov-Kette heißt *aperiodisch*, falls $d = 1$.

Satz 48 (Döblin)**.** *Eine (X_k) eine positiv rekurrente, irreduzible und aperiodische Markovkette ist mischend, d.h. für eine beliebige Startverteilung λ gilt*

$$P_\lambda \left(X_k = j \right) \stackrel{j \to \infty}{\longrightarrow} \pi(j),$$

wobei π die eindeutig bestimmte invariante Verteilung von P ist.

Beispiel 49 (Gästeverteilung bei einer Party)**.** Wir feiern eine Party in unserer Wohnung und laden dazu unsere Freunde ein. Jeder Gast bewegt sich dabei unabhängig von den anderen Gäste enstprechend einer (δ_{Diele}, P)-Markov Kette, wobei die Übergangsmatrix P die Wahrscheinlichkeiten der Wechsel zwischen den Zimmern beschreibt. Wir nehmen an, dass die Übergangsmatrix aperiodisch und irreduzibel ist. Nach langer Zeit ergibt sich dann für die Anzahl der z.B. in der Küche anzutreffenden Gäste, dass

$$\frac{\text{Anzahl der Personen in Küche}}{\text{Anzahl der Personen im Haus}} \approx \frac{1}{\mathbb{E}_{\text{Küche}} \left(\tau_{\text{Küche}} \right)}$$

Für den Beweis verwenden wir das folgende Lemma aus der elementaren Zahlentheorie.

? **Lemma 50.** *Falls $M \subset \mathbb{N}$ abgeschlossen bzgl. Addition und $ggT(M) = 1$, dann existiert $m \in \mathbb{N}$, s.d. $\mathbb{N}_{\geq m} \subset M$.*

Beweis. 1.Schritt: Zeige, dass $\exists N, N+1 \in M$. Sei $n_0 \in M, n_0 + k = n_1 \in M$. Falls $k = 1$ klar. Falls $k > 1$: $ggT = 1; k > 1$. Dann folgt $l \in M$ und $k \nmid l$.

$$l = mk + \sigma, \, \sigma \in (0, k)$$

$$\begin{aligned} \tilde{n}_0 &:= (m+1)n_0 + l \in M \\ \tilde{n}_1 &:= (m+1)(n_0 + k) \in M \end{aligned}$$

weil $n_0 \in M$, s.d. $(m+1)n_0 \in M$, s.d. $(m+1)n_0 + l \in M$ ebenso mit $(n_0 + k) = n_1 \in M$

$$\begin{aligned} \tilde{n}_1 - \tilde{n}_0 &= k(m+1) - l \\ &= (k - \sigma), \, \sigma \geq 1 \end{aligned}$$

Dann ist

$$\tilde{n}_1 - \tilde{n}_0 < n_1 - n_0$$

Nach ~~endl.~~ vielen Iterationsschritten folgt Behauptung.

endlich

Schritt 2: Sei $m_0 := N^2$ und $m \geq m_0$

$$m - N^2 = kN + \sigma, \sigma \in [0, N)$$

Dann

$$\begin{aligned} m &= N^2 + \sigma + kN \\ &= \sigma \underbrace{(1 + N)}_{\in M} + (N - \sigma + k) \underbrace{N}_{\in M} \in M \end{aligned}$$

$\square$

Beweis von Satz 48. Wir verwenden ein sogenanntes *Kopplungsargument*, welches darin besteht, zwei Prozesse Y und Z zu konstruieren, die von unterschiedlichen Anfangsbedingungen starten, die für sich allein betrachtet jeweils Markov-Ketten mit vorgegbenen Übergangskernen aber nicht unabhängig sind, sondern fast sicher in endlicher Zeit aneinander koppeln. Zur Realisierung dieser Koppelung sei neben der (λ, P)-Markovkette (X_k) zunächst noch eine weitere von (X_k) unabhängige (π, P)-Markovkette (Y_k) konstruiert. Wir definieren hiermit den stochastischen Prozess $W_k := (X_k, Y_k) \in E \times E$, der eine Markovkette auf $E \times E$ darstellt mit Übergangsmatrix

$$\begin{aligned} Q_{ii',jj'} &= P_{ij} \cdot P_{i'j'} \\ &= P\left(W_{k+1} = (j, j') \mid W_k = (i, i')\right) \\ &= P\left(X_{k+1} = j, Y_{k+1} = j' \mid X_k = i, Y_k = i'\right) \\ &= P\left(X_{k+1} = j \mid X_k = i\right) \cdot P\left(Y_{k+1} = j' \mid Y_k = i'\right) \end{aligned}$$

Die Startverteilung von (W_k) ist dann $\lambda \otimes \pi$, denn $P\left(W_0(i, j)\right) = \lambda(i)\pi(j) = (\lambda \otimes \pi)[i, j]$.

e (Produktmaß)

Wir behaupten nun, dass Q auf $E \times E$ irrduzibel ist. Dazu stellen wir zunächst fest, dass für $i, j \in E$ ein $m = m(i, j)$ existiert mit

$$P_{ij}^{(n)} > 0 \forall n \geq m,$$

denn die Menge $M := \left\{ l \mid P_{ij}^{(l)} > 0 \right\}$ erfüllt aufgrund der Aperiodzität von P die Voraussetzungen von Lemma 50. Hieraus folgt die Irreduzibilität von Q, denn für $(i, i'), (j, j')$ folgt mit $m = \max\left(m(i, j), m(i', j')\right)$:

$$Q_{ii',jj'}^{(m)} = P_{ij}^{(m)} P_{i'j'}^{(m)} > 0$$

Zudem ist Q ist positiv rekurrent, denn $\pi \otimes \pi$ ist als W-Maß auf $E \times E$ eine invariante Verteilung für Q. Dann gilt unter $P_{\lambda \otimes \pi}$:

$$P_{\lambda \otimes \pi}\left(\tau_{(u,v)} < \infty\right) = 1$$

d.h. jeder Zustand (u, v) in $E \times E$ wird von $(W_k)_k$ fast sicher in endlicher Zeit erreicht. Wählen wir z.B. $(u, v) = (b, b) \in diag(E \times E)$, so gilt

$$P_{\lambda \otimes \pi}\left(\tau_{(b,b)} < \infty\right) = 1.$$

Schließlich definieren wir einen neuen Prozess $(Z_k)_{k \geq 0}$ auf E durch

$$Z_k := \begin{cases} X_k & \text{falls } k < \tau_{(b,b)} \\ Y_k & \text{falls } k \geq \tau_{(b,b).} \end{cases}$$

Bildlich gesprochen beschreibt (Z_k) die Reise eines Passagiers, der zunächst mit dem Prozess (X_k) mitreist und bei der ersten Begegnung von X und Y entscheidet, seine Reise künftig mit Y fortzusetzen. Insbesondere ist Z ebenfalls im Sinne der Markov-Eigenschaft gedächtnislos, und die Übergangsmatrix ist erneut P. Da Z gemeinsam mit X startet, ist die Anfangsverteilung ebenfalls λ, d.h. $(Z_k)_k$ also eine (λ, P)-Markovkette, hat also insbesondere dieselben Verteilungen wie (X_k). Daher

$$\begin{aligned}
|P_\lambda(X_n = k) - \pi(k)| &= |P_\lambda(X_n = k) - P_\pi(Y_n = k)| \\
&= |P_\lambda(Z_n = k) - P_\pi(Y_n = k)| \\
&= |P_{\lambda\otimes\pi}(Z_n = k) - P_{\lambda\otimes\pi}(Y_n = k)| \\
&= |P_{\lambda\otimes\pi}(Z_n = k; n \leq \tau) + \underbrace{P_{\lambda\otimes\pi}(Z_n = k; n \geq \tau)}_{=P(Y_n = k; n \geq \tau)} \\
&\qquad - P(Y_n = k; n \leq \tau) - P(Y_n = k; n \geq \tau)|
\end{aligned}$$

~~und~~

$$\begin{aligned}
\cancel{|P(Z_n = k) - P(Y_n = k)|} &\cancel{= |P(Z_n = k; n < \tau) + P(Z_n = k; n \geq \tau)} \\
&\qquad \cancel{- P(Y_n = k; n < \tau) - P(Y_n = k; n \geq \tau)|} \\
&= |P(Z_n = k; n < \tau) - P(Y_n = k; n < \tau)| \\
&\leq P(n < \tau), n \to \infty \\
&\longrightarrow 0, \text{ weil } P(\tau = \infty) = 0
\end{aligned}$$

Bem.: Satz von Perron-Frobenius

□

Bemerkung 51. Für eine Verallgemeinerung des Döblin'schen Satzes auf periodische Markovketten konsultiere man z.B. das Lehrbuch 'Markov Chains' von J. Norris, Cambridge University Press.

2. Der Birkhoffsche Ergodensatz

Wir schließen das Kapitel über das Langzeitverhalten von stochastischen Prozessen mit einen fundamentalen Resultat, welches nicht nur in der Wahrscheinhlichkeitstheorie sondern z.B. bei dynamischen Systemen von großer Bedeutung ist. Die Entstehung dieses Resultates geht auf die Frage in der theoretischen Physik zurück, ob ein mechanisches Mehrteilchensystem im zeitlichen Mittel gegen ein 'statistisches Gleichgewicht' strebt, wobei letzteres aufzufassen ist als eine Wahrscheinlichkeitsverteilung auf der Menge der möglichen Zustände des Systems. Eine sehr allgemeine Antwort auf diese Frage liefert der Birkhoff'sche Ergodensatz von 1931. Er ist eine sehr mächtige abstrakte Erweiterung des starken Gesetzes der großen Zahlen.
Als mathematischer Satz handelt er von sogenannten *maßerhaltenden Transformationen* eines Wahrscheinlichkeitsraumes. Vor der allgemeinen Definition betrachten wir zunächst das Beispiel der Zeitverschiebung bei stationären Prozessen.

2.1. Maßerhaltende Transformationen.

Definition 52. Ein stochastischer Prozess $(X_n)_{n\in\mathbb{N}}$ mit messbaren Zustandsraum $(E, \mathcal{E})$ heißt *stationär*, wenn die Verteilungen von $(X_k)_{k\in\mathbb{N}}$ und $(X_{k+1})_{k\in\mathbb{N}}$

gleich sind, d.h. wenn für alle $l \in \mathbb{N}$ und alle messbare Mengen $A_1, \cdots, A_l \subset E$ gilt, dass

$$\begin{aligned} P(X_1 \in A_1, ..., X_l \in A_l) &= P(X_2 \in A_1, ..., X_{l+1} \in A_l) \\ &= P(X_k \in A_1, ..., X_{k+l} \in A_l) \end{aligned}$$

Die Wahrscheinlichkeiten für das Eintreten von Szenarien von Trajektorien sind also unabhängig von der Festlegung des Startzeitpunktes $t_0 = 0$.

Beispiel 53. (1) Eine Folge von identisch unabhängigen Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ is stationär.

(2) Es sei $(X_k)_{k\in\mathbb{N}_0}$ eine (π, P)-Markovkette auf dem diskreten Zustandsraum E, wobei π eine invariante Verteilung von P ist. Dann ist (X_k) stationär (siehe Übungen).

Wir sagen, dass die Verteilungen von (X_k) in diesem Falle *Shift-invariant* sind, wobei 'Shift' hier den Zeit-Shift, d.h. die Verschiebung des Startzeitpunktes bezeichnet. Um diese Eigenschaft in einen allgemeineren Kontext zu bringen, führen wir nun noch einige weitere nützliche Sprechweisen ein.

Definition 54. Es sei $(E, \mathcal{E})$ messbarer Raum und I eine (Zeit-)Indexmenge, dann heißt

$$\hat{\Omega} = E^I = \{\omega = (\omega_i)_{i\in I}\}$$

der *kanonische Pfadraum.* Zu $i \in I$ heißt

$$\pi_i : \hat{\Omega} \to E, \quad (\omega_j)_{j\in I} \to \omega_i,$$

die *Koordinatenabbildung zur Koordinate* $i \in I$.
$\hat{\Omega}$ kann mit der natürlichen Produkt-sigma-Algebra

$$\hat{\mathcal{F}} := \mathcal{E}^{\otimes I} = \sigma\{\pi_i, i \in I\}$$

ausgestattet werden.

Definition 55. Ein stochastischer Prozess $(X_i)_{i\in I}$, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und mit Zustandsraum E induziert vermöge der *Pfadabbildung* $(E, \mathcal{E})$

$$X_\cdot : \Omega \to \hat{\Omega} = E^I, \quad X_\cdot(\omega) = (X_i(\omega))_{i\in I} \in E^I$$

ein Wahrscheinlichkeitsmaß auf $(\hat{\Omega}, \hat{\mathcal{F}})$ als Bildmaß von P unter X.

$$P_{X_\cdot}(A) = P(X_\cdot \in A), \quad A \in \hat{\mathcal{A}}$$

welches *Verteilung von* $(X_i)_{i\in I}$ genannt wird.

Im Fall, dass $I = \mathbb{N}$ (bzw. $I = \mathbb{N}_0$) können wir schließlich den sogenannten *Zeitshiftoperator* auf dem Pfadraum definieren wie folgt

$$\Theta : E^{\mathbb{N}} \to E^{\mathbb{N}}, \quad \Theta[(\omega_n)_{n\in\mathbb{N}}] := (\omega_{n+1})_{n\in\mathbb{N}}.$$

Der Zeitshift ist messbar als Abbildung des messbaren Raums $(\hat{\Omega}, \hat{\mathcal{F}})$ in sich selbst, so dass folgene Definition möglich ist.

Definition 56. Ein Maß μ auf $(\hat{\Omega} = E^{\mathbb{N}}, \hat{\mathcal{F}})$ heißt *Shift-invariant*, falls $\Theta_*\mu = \mu$.

In dieser Sprechweise ist ein stochastischer Prozess $(X_k)_{k\in\mathbb{N}}$ somit stationär genau dann, wenn seine Verteilung $P_{X.}$ auf dem kanonischen Pfadraum $\hat{\Omega} = E^{\mathbb{N}}$ Shift-invariant ist.

Bemerkung 57. Der auf dem kanonischen Pfadraum $\hat{\Omega}$ definierte Prozess

$$\hat{X}_k : \Omega \to E, \quad \hat{X}_k(\omega) := \pi_k(\omega) = \omega_k$$

wird auch als *(kanonischer) Koordinatenprozess* bezeichnet. Wir werden im Abschnitt über zeitstetige Markovprozesse (wenn $I = \mathbb{R}_{\geq 0}$) darauf zürückkommen.

Wir können den Fall eines stationären Prozesses nun als Spezialfall einer viel allgemeineren Situation auffassen.

Definition 58 (Maßerhaltende Transformation)**.** Sei $(\Omega, \mathcal{J}, P)$ ein Wahrscheinlichkeitsraum, $T : \Omega \to \Omega$ messbar. T heißt maßerhaltend falls

$$T_*P = P.$$

Beispiel 59 (Rotation der Einheitssphäre)**.** Sei $\Theta \in [0, 2\pi)$ ein fest gewählter (Dreh-)Winkel.

$$\Omega = S^1 = \{x \in \mathbb{C} \mid |x| = 1\}$$
$$\mathcal{F} = \mathcal{B}\left(S^1\right)$$
$$P\left([\alpha, \beta)\right) = \frac{\beta - \alpha}{2\pi}$$
$$T_\Theta : S^1 \to S^1;\ z \mapsto e^{i\Theta} z$$

Dann ist T_Θ ist maßerhaltend.

Es sei $(\Omega, \mathcal{A})$ ein messbarer Raum und P ein W-Maß auf $(\Omega, \mathcal{A})$. Dann...

Definition 60. Eine Menge $A \subset \Omega$ heißt T-*invariant* falls

$$T(A) = A$$

Die T-*invariante σ-Algebra* ist definiert als

$$\mathcal{I}_T := \{A \in \mathcal{F} \mid T(A) = A\}$$

Eine P-

$T : \Omega \to \Omega$ heißt *ergodisch*, falls $\mathcal{I}_T$ P-trivial ist, d.h.

$$P(A) \in \{0, 1\} \quad \forall A \in \mathcal{I}_T.$$

Beispiel 61.

(1) Der Zeitshift auf dem kanonischen Pfadraum ist ergodisch für die Verteilung einer i.i.d.Folge von Zufallsvariablen, denn falls

$$A = \Theta A = \Theta^2 A = \Theta^m A$$

gilt, dann folgt

$$A \in \mathcal{T}$$

mit der Terminalen σ-Algebra $\mathcal{T}$ (siehe Vorlesung Wahrscheinhlichkeitstheorie I). Mit dem Kolmogorov'schen 0-1-Gesetz folgt

$$P(A) \in \{0, 1\}$$

(2) $T = T_\Theta$ ist als Rotation des Einheitskreises um den Winkel $\Theta \in [0, 2\pi)$ ergodisch genau dann, wenn

$$\Theta \notin \left\{ \frac{2\pi}{k}, k \in \mathbb{N} \right\}.$$

Wir skizzieren hier eine Begründung (einen exakten alternativen Beweis mit Fourier-Analysis findet man etwa im Skript *Wahrsheinlichkeitstheorie* von Wolfgang König). Sei $\Theta = \frac{2\pi}{k}$ mit $k \in \mathbb{N}$, s.d. $(T_\Theta)^k z = z$ für alle z, $Te^k = id$. Sei $A \in \mathcal{F}$, $P(A) = f > 0$ mit A klein, d.h. in einem geeigneten kleinen Intervall der Spähre enthalten. Dann sind die Mengen $\{T^j(A)\}_{j \in [0,k-1]}$ disjunkt. Die Menge

$$\overline{A} := \bigcup_{j=1}^{k-1} T^j(A)$$

ist dann T invariant mit

$$1 > P(\overline{A}) = k \cdot f > 0$$

falls $f < \frac{1}{k}$. Somit ist T nicht ergodisch. – Falls $\Theta \notin \left\{\frac{2\pi}{k}; k \in \mathbb{N}\right\}$ ist, dann definiert für $z \in S^1$

$$R := \{T^n(z), n \in \mathbb{N}\}$$

eine diskrete, aperiodische Menge, d.h. für für $r \in R$ gilt $T^l(r) = r \Leftrightarrow l = 0$. Insbesondere ist R eine dichte Teilmenge in S^1. Sei nun beispielsweise $A \subset S^1$ abgeschlossen, nichtleer T-invariant, so folgt mit $z \in A$

$$\exists z \in A \Longrightarrow z_1 := T_\Theta(z) \in A, \ldots \forall T^k(z) \in A \Longrightarrow \underbrace{\overline{R}}_{=S^1} \subset A \Rightarrow \lambda(A) = 2\pi.$$

Für allgemeine T-invariante messbare Mengen A ergibt sich analog durch Approximation

$$\lambda(A) \in \{0, 2\pi\}.$$

integrierbare numerische

Also ist T in diesem Fall ergodisch.

2.2. **Wiederholung: Bedingte Erwartung.** Für die Formulierung des Birkhoffschen Ergodensatzes erinnern wir noch einmal an die bedingte Erwartung einer Zufallsvariable, die wir in der Wahrscheinhlichkeitstheorie I eingeführt haben.

Definition 62. Es sei $X : \Omega \to \overline{\mathbb{R}}$ eine ~~reelle~~ Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und $\mathcal{G} \subset \mathcal{F}$ eine Unter-σ-Algebra. Dann heißt eine reelle Zufallsvariable Y auf Ω *bedingte Erwartung (unter P) von X gegeben $\mathcal{G}$*, fall Y messbare bzgl. $\mathcal{G}$ ist und für alle $G \in \mathcal{G}$ gilt

$$\int_G X dP = \int_G Y dP.$$

Lemma 63. *Die bedingte Erwartung ist fast sicher eindeutig bestimmt.*

Beweis. Denn für zwei bedingte Erwartungen Y_1 und Y_2 gilt, dass $A := \{Y_1 > Y_2\} \in \mathcal{G}$. Also gilt $\int_A Y_1 dP = \int_A X dP = \int_A Y_2 dP$, also $\int_{Y_1 > Y_2} (Y_1 - Y_2) dP = 0$, was nur möglich ist, wenn A eine P-Nullmenge ist. $\square$

Wir dürfen also von *der* bedingten Erwartung sprechen und schreiben $Y = E_P(X|\mathcal{G})$.

Der Name *bedingte Erwartung* ist u.A. dadurch gerechtfertigt, dass im Fall einer endlich disjunkt erzeugten σ-Algebra

$$\mathcal{G} = \sigma(A_1, \cdots, A_N), \quad \text{mit } A_i \in \mathcal{F}, \quad \bigcup_{i=1}^N A_i = \Omega, \quad \mathcal{A}_i \cap A_j = \emptyset \text{ für } i \neq j,$$

und $P(A_i) > 0$ für alle A_i gilt, dass

$$E_P(X|\mathcal{G})(\omega) = \sum_{i=1}^N E_P(X|A_i)P(A_i)1_{A_i}(\omega),$$

wobei

$$E_P(X|A_i) = \frac{E_P(X;A_i)}{P(A_i)}$$

den *bedingten Erwartgungswert von X gegeben das Ereignis A_i* bezeichnet. Als abstraktes Existenzresultat haben wir das folgende.

Satz 64. *Falls in der Situation von Definition 62 die Zufalsvariable X P-integrierbar ist, so existiert $E_P(X|\mathcal{G})$.*

Beweis. Man gehe zunächst davon aus, dass $X \in L^2((\Omega, \mathcal{F}, P)$. Die Menge $L^2(\Omega, \mathcal{G}, P) \subset L^2(\Omega, \mathcal{F}, P)$ stellt dann eine gegenüber dem L^2-Abstand auf $L^2(\Omega, \mathcal{F}, P)$ abgeschlossene Teilmenge dar, und $Y = E_P(X|\mathcal{G})$ kann gewonnen werden als orthogonale Projektion von X in L^2 auf die Teilmenge $L^2(\Omega, \mathcal{F}, P)$, d.h. Y ist der Minimierer von

$$\inf_{Z \ \in L^2(\Omega,\mathcal{G},P)} \|Z - X\|^2_{L^2(\Omega,\mathcal{F},P)}.$$

Der allgemeinere Fall von integrierbarem X ergibt sich hieraus durch Approximation. (Details im Skript zur Wahrscheinlichkeitstheorie I). $\square$

Die Operation der bedingten Erwartungen hat ganz analoge Eigenschaften wie die konventionelle Erwartungswertswertbildung. Wir geben eine Auswahl der wichtigtsten davon an.

Satz 65 (Eigenschaften der bedingten Erwartung)**.**

(1) *$E(\lambda X + \mu Y|\mathcal{G}) = \lambda E(\lambda X|\mathcal{G}) + \mu E(Y|\mathcal{G})$ (Linearität)*
(2) *$X \geq 0$ f.s. $\Rightarrow E(X|\mathcal{G}) \geq 0$ f.s. (Positivität)*
(3) *Für $\mathcal{H} \subset \mathcal{G} \subset \mathcal{G}$ ist $E(X|\mathcal{H}) = E(E(X|\mathcal{G}n)|\mathcal{H})$ ('Turmeigenschaft')*
(4) *Für $\varphi : \mathbb{R} \to \mathbb{R}$ convex $E(\varphi(X)|\mathcal{G}) \leq \varphi(E(X|\mathcal{G})$. (Jensen'sche Ungleichung)*
(5) *Für $X, Y \in L^p((\Omega, \mathcal{F}, P)$, $p \geq 1$, ist $\|E(X|\mathcal{G}) - E(\mathcal{Y}|\mathcal{G})\|_{L^p} \leq \|X - Y\|_{L^p}$ (Stetigkeit in L^p).*
(6) *Falls $\sigma(X)$ und $\mathcal{G}$ unabhänigig gilt $E(X|\mathcal{G}) = E(X)$*
(7) *Für Y beliebig und X messbar bzgl. $\mathcal{G}$ ist $E(X \cdot Y|\mathcal{G}) = X \cdot E(Y|\mathcal{G})$.*

Beweis. Die Eigenschaften (1) – (3) und (7) lassen sich direkt aus Definition 62 und der fast sicheren Eindeutigkeit der bedingten Erwartung ableiten, ebenso wie die Aussage (6) im Spezialfall, dass X die Indikatorfunktion eines von $\mathcal{G}$ unabhängigen Ereignisses ist. Durch Approximation mit elementaren Funktionen gewinnt man hieraus die allgemeine Aussage.

Für die Jensen'sche Ungleichung (4) erinnern wir uns, dass der Wert einer konvexen Funktion als punktweise gebildetes Supremum sämtlicher affin-linearen unteren Stützgeraden gebildet werden kann, d.h. für alle $x \in \mathbb{R}$ gilt

$$\varphi(x) := \sup_{(m,c)\in Z} m \cdot x + c \tag{3}$$

wobei die Menge

$$Z = \{(m,c) \in \mathbb{R} \times \mathbb{R} | m \cdot x + c \leq \varphi(x) \, \forall x \in \mathbb{R}\}$$

von allen zulässigen Paare von Steigungen und Achsverschiebungen gebildet wird, für die die zugehörigen affin-linearen Funktionen unterhalb des Graphen von φ liegen. Insbesondere gitl für $(m,c) \in Z$

$$\varphi(X) \geq mX + c \text{ f.s. },$$

was zusammen mit den Eigenschaften (1) und (2) liefert, dass

$$E(\varphi(X)|\mathcal{G}) \geq mE(X|\mathcal{G})(\omega) + c \text{ f.s.}$$

Die Darstellung (3) von φ bleibt richtig, wenn wir das surpremum auf der rechten Seite nur über die abzählbare Teilmenge $Z' := Z \cap \mathbb{Q} \times \mathbb{Q}$ bilden. Hiermit gilt dann nach Auswahl einer geeigneten Nullmenge für P-fast alle $\omega \in \Omega$, dass

$$E(\varphi(X)|\mathcal{G})(\omega) \geq mE(X|\mathcal{G})(\omega) + c \quad \forall (m,c) \in Z',$$

und die behauptete Jensen'sche Ungleichung folgt durch Übergang zum Supremum auf der rechten Seite bzgl. $(m,c) \in Z'$. Aus der Jensen'schen Ungleichung folgt mit $\varphi(x) = |x|^p$ auch die Eigenschaft (5), denn

$$\begin{aligned}\|E(X|\mathcal{G}) - E(Y|\mathcal{G})\|_{L^p}^p &= E(|E(X-Y|\mathcal{G})|^p) \\ &\leq E(E(|X-Y|^p|\mathcal{G})) = E(|X-Y|^p) = \|X-Y\|_{L^p}^p.\end{aligned}$$

□

2.3. **Formulierung und Beweis vom Birkhoffschen Ergodensatz.**

Satz 66. *Sei $(\Omega, \mathcal{F}, P)$ ein Maßraum und $\tau : \Omega \to \Omega$ eine maßerhaltende Transformation, $f : \Omega \to \mathbb{R} \in L^1(\Omega, \mathcal{F}, P)$. Dann gillt*

$$\frac{1}{n}\sum_{j=0}^{n-1} f\left(\tau^{(j)}(\omega)\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}_P\left(f \mid \mathcal{I}_\tau\right)(\omega) \text{ für } P\text{-f.a. } \omega \in \Omega.$$

Bemerkung 67. Falls τ ergodisch ist, gilt $E_P(f|\mathcal{I}_T) = E_P(f)$ fast sicher, d.h. der Grenzwert ist dann deterministisch.

Beweis. Wir geben einen Beweis in zwei Schritten.
Schritt 1: Zeige

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{j=0}^{n-1} f\left(\tau^{(j)}(\omega)\right) =: S_n(\omega)\right) =: Z \text{ existiert fast sicher.}$$

Schritt 2: Zeige

$$Z = \mathbb{E}_P\left(f \mid \mathcal{I}_\tau\right).$$

Wir verwenden im folgenden die probabilistische Schreibweise X für die reelle Zufallsvariable $X := f : \Omega \to \mathbb{R}$.

Zu 1) Durch Zerlegung in Positiv- und Negativteil von X können wir o.B.d.A. davon ausgehen, dass $X \geq 0$. Wir führen die Zufallsvariablen

$$\overline{X} := \limsup_{k\to\infty} S_k; \underline{X} := \liminf_{k\to\infty} S_k$$

ein und wollen zeigen, dass

$$E\overline{X} \leq EX \text{ bzw. } E\underline{X} \geq EX,$$

also insbesondere $E\overline{X} \leq E\underline{X}$, was wegen $\overline{X} \geq \underline{X}$ f.s. bedeutet, dass

$$\overline{X} = \underline{X} \text{ f.s.}$$

und

$$\mathbb{E}(Z) = \mathbb{E}(X) = \mathbb{E}\left(\underline{X}\right) = \mathbb{E}\left(\overline{X}\right).$$

Im folgenden führen wir den Beweis für die Aussage

$$\int \overline{X} dP \leq \int X dP,$$

der Beweis für $\int \underline{X} dP \geq \int X dP$ verläuft analog. – Sei hierzu $M \geq 0$, $\overline{X}_M := \overline{X} \wedge M, \overline{X}$. Beachte, dass $\overline{X}$ τ-invariant ist und daher ebenso $\overline{X}_M$, d.h. also

$$\overline{X}(\tau(\omega)) = \overline{X}(\omega) \quad \forall \omega \in \Omega.$$

Zu $\epsilon > 0$ sei

$$n := \inf\left\{k \mid S_k(\omega) \geq \overline{X}_M - \varepsilon\right\}$$

Dann ist n fast sicher endlich, folglich existiert ein $N_\epsilon \in \mathbb{N}$, so dass

$$P\left(n > N_\varepsilon\right) \leq \varepsilon.$$

Wir führen die Ausnahmemenge

$$A := \{\omega \in \Omega \mid n(\omega) > N_\varepsilon\}$$

ein und setzen

$$\begin{aligned}\tilde{X}(\omega) : &= X(\omega) 1_{A^c}(\omega) + 1_A(\omega) M \\ \tilde{n}(\omega) : &= n(\omega) 1_{A^c}(\omega) + 1_A(\omega)\end{aligned}$$

Dann gilt für alle $\omega \in \Omega$

$$\begin{aligned}1)\ \tilde{X}(\omega) &\geq X(\omega) \\ 2)\ \frac{1}{\tilde{n}(\omega)} \sum_{j=0}^{\tilde{n}(\omega)-1} \tilde{X}\left(\tau^{(j)}(\omega)\right) &\geq \overline{X}_M(\omega) - \varepsilon.\end{aligned}$$

Definiere nun iterativ:

$$n_0 := 0;\ n_1 := \tilde{n};\ n_k := n_{k-1} + \tilde{n}\left(\tau^{n_{k-1}}(\omega)\right)$$

Für $m \in \mathbb{N}$ sei

$$k_m(\omega) := \max\left\{k \mid n_k(\omega) \leq m\right\}.$$

Dann gilt wegen $\tilde{n} \leq N_\epsilon$ stets

$$m - n_{k_m} \leq N_\varepsilon$$

$$
\begin{aligned}
\sum_{j=0}^{m-1} \tilde{X}\left(\tau^{(j)}(\omega)\right) &\geq \sum_{j=0}^{n_{k_m}-1} \tilde{X}\left(\tau^{(j)}(\omega)\right) \\
&= \sum_{j=0}^{n_1-1} \tilde{X}\left(\tau^{(j)}(\omega)\right) + \sum_{j=n_1}^{n_2-1} \tilde{X}\left(\tau^{(j)}(\omega)\right) + \ldots + \sum_{j=n_{k_m-1}}^{n_{k_m}-1} \tilde{X}\left(\tau^{(j)}(\omega)\right) \\
&\geq n_1(\omega)\left[\overline{X}_M(\omega) - \varepsilon\right] + \left[n_2(\omega) - n_1(\omega)\right]\left[\overline{X}_M\left(\tau^{n_1}(\omega)\right) - \varepsilon\right] + \ldots \\
&\quad + \left[n_{k_m} - n_{k_m-1}\right](\omega)\left[\overline{X}_M\left(\tau^{n_{k_m-1}}(\omega)\right) - \varepsilon\right] \\
&= n_1(\omega)\left[\overline{X}_M(\omega) - \varepsilon\right] + [\ldots]\left[\overline{X}_M(\omega) - \varepsilon\right] + \ldots\left[\overline{X}_M(\omega) - \varepsilon\right] \\
&= n_{k_m}(\omega)\left[\overline{X}_M(\omega) - \varepsilon\right] \\
&= m\left[\overline{X}_M(\omega)\right] + \left[n_{k_m} - m\right]\left[\overline{X}_M(\omega)\right] - n_k\varepsilon
\end{aligned}
$$

Wir schätzen nun ab

$$
\begin{aligned}
\int \tilde{X} dP &= \int_{A^c} X dP + \int_A M dP \\
&\leq \int X dP + M\varepsilon,
\end{aligned}
$$

so dass

$$
\begin{aligned}
\int X dP &\geq \int \tilde{X} dP - M\varepsilon \\
&= \frac{1}{m}\int \sum_{j=0}^{m-1} \tilde{X}\left(\tau^{(j)}(\omega)\right) P(d\omega) - M\varepsilon \\
&\geq \int \overline{X}_M(\omega) P(d\omega) + \frac{\overbrace{n_{k_m} - m}^{\leq N_\varepsilon}}{m} \int \underbrace{\overline{X}_M(\omega)}_{|\ldots| \leq M} - \underbrace{\frac{n_{k_m}}{m}}_{\leq 1}\varepsilon - M\varepsilon \\
&\geq \int \overline{X}_M dP - \frac{N\varepsilon}{m} M - \varepsilon - M\varepsilon \\
&\overset{m\to\infty}{\longrightarrow} \int \ldots - \varepsilon - M\varepsilon \\
&\overset{\varepsilon\to 0}{\longrightarrow} \int \overline{X}_M dP
\end{aligned}
$$

Für $M \to \infty$ folgt schließlich wie behauptet, dass

$$
\int X dP \geq \int \overline{X} dP
$$

Zum zweiten Schritt. Wir wollen nun zeigen, dass

$$
Z = \mathbb{E}_P\left(X \mid \mathcal{J}\right)
$$

ist zz. Dazu sei $A \in \mathcal{J}$. Definieren wir $Y := 1_A X$, so folgt gemäß dem im ersten Schritt gezeigten, dass

$$
\tilde{S}_n := \frac{1}{n}\sum_{j=0}^{n-1} Y\left(\tau^{(j)}(\omega)\right) \overset{n\to\infty}{\longrightarrow} \tilde{Z}
$$

mit $\mathbb{E}_P\left(\tilde{Z}\right) = \mathbb{E}_P(Y) = \int_A X dP$. Ferner gilt unter Ausnutzung der τ-Invarianz von A, dass

$$\begin{aligned}\tilde{Z} &= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} 1_A\left(\tau^{(j)}(\omega)\right)\cdot X\left(\tau^{(j)}(\omega)\right)\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} 1_A(\omega) X\left(\tau^{(j)}(\omega)\right)\\ &= 1_A(\omega)\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} X\left(\tau^{(j)}(\omega)\right)\\ &= 1_A Z\end{aligned}$$

Also ist

$$\tilde{Z} = 1_A Z \text{ fast sicher.}$$

und somit gilt gemäß dem in Schritt gezeigten, dass

$$\begin{aligned}\mathbb{E}_P\left(\tilde{Z}\right) &= \int \tilde{Z} dP\\ &= \int_A Z dP = E(Y) = \int_A X dP.\end{aligned}$$

Zudem ist

$$Z = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} X\left(\tau^{(j)}(\omega)\right)$$

$\tau-$invariant, also $\mathcal{J}-$messbar, woraus insgesamt folgt, dass

$$Z = \mathbb{E}_P\left(X \mid \mathcal{J}\right).$$

□

Beispiel 68. (1) Starkes Gesetz der Großen Zahlen:
Falls $(X_k)_{k\in\mathbb{N}}$ iid. ist mit $\mathbb{E}\left(|X_1|\right) < \infty$, dann gilt

$$\frac{1}{n}\sum_{k=1}^{n} X_k(\omega) \overset{n\to\infty}{\longrightarrow} \mathbb{E}(X_1) \text{ fast sicher.}$$

Die Aussage kennen wir schon aus der W-Theorie I, wir wollen sie hier aus dem Ergodensatz folgern. In der Tat, mit

$$(\Omega, \mathcal{J}, P) = \left(\mathbb{R}^N, \mathcal{B}\left(\mathbb{R}^N\right), \bigotimes_{j=1}^{\infty} \mathrm{Vert}\left(X_1\right)\right)$$

(unendlicher Produkraum, siehe W-Theorie I) ist der Zeitshift

$$\tau : (\omega_1, ...) \to (\omega_2, ...)$$

eine maßerhaltende Transformation. Nach dem Kolmogrov'schen 0-1-Gesetz ist τ ergodisch, also folgt mit Birkhoff

$$\Xi : \Omega \longrightarrow \mathbb{R},\ (\omega_1, ...) \mapsto \omega_1$$

$$\begin{aligned} \Longrightarrow \Xi\left(\tau^{(k)}(\omega)\right) &= \omega_k \\ \Longrightarrow \frac{1}{n}\sum_{j=1}^{k}\Xi\left(\tau^{k}(\omega)\right) &= \frac{1}{n}\sum_{j=1}^{n}\omega_j \end{aligned}$$

Dann:

$$\frac{1}{n}\sum_{j=1}^{n}\Xi\left(\tau^{(k)}(\omega)\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}_P\left(\Xi \mid \mathcal{J}\right) = \mathbb{E}_P\left(\Xi\right) = \mathbb{E}\left(X_1\right)$$

für P-f. alle ω.

(2) Analog ergibt sich unmittelbar die folgende Erweiterung der starken Gesetzes. In der Situation wie oben gilt für jede (beschränkte) messbare Funktion $h : \mathbb{R}^d \to \mathbb{R}$, dass

$$\frac{1}{n}\sum_{k=1}^{n} h(X_{k+1}, \cdots, X_{k+d}) \overset{n\to\infty}{\longrightarrow} E(h(X_1, \cdots, X_d)) \text{ fast sicher.}$$

(3) Wir betrachten eine Markovkette in fünf Zuständen a, b, c, d, e mit:

$$\begin{aligned} a &\to b, d \\ b &\to b, c \\ c &\to b, c \\ d &\to d, e \\ e &\to d, e \end{aligned}$$

Dann gibt es zwei inv. Verteilungen, die jeweils links und rechts konzentriert sind. Sie seien mit λ und μ bezeichnet. λ und μ sind zudem ergodisch. Für alle $\alpha \in [0,1]$ ist dann

$$P_\alpha = \alpha\mu + (1-\alpha)\lambda$$

ebenfalls invariant aber für $\alpha \in]0,1[$ nicht ergodisch. Der Ergodensatz ist auf (P_α, τ) anwendbar z.B. auf die Zufallsvariable

$$X(\omega) = 1_{\{\omega_1 \in b\}} + 1_{\{\omega_1 \in d\}}$$

mit $(\omega_1, ...,)$ Pfad als Realisierung von (X_k). Mit Birkhoff gilt nun, dass unter P_α

$$\frac{1}{n}\sum_{j=0}^{n-1} X\left(\tau^j(\omega)\right) \longrightarrow \mathbb{E}_{P_\alpha}\left(X \mid \mathcal{J}\right) \text{ fast sicher.}$$

Die Zufallsvariable $\mathbb{E}_{P_\alpha}\left(X \mid \mathcal{J}\right)$ kann explizit berechnet werden (vergl. Übungen).

3. Martingale

3.1. **Grundlegende Eigenschaften.** Im Folgenden betrachten wir stochastische Prozesse $(X_k)_{k\in I}$, deren Parametermenge I eine Teilmenge von $\mathbb{R}$ ist. (In der Tat lassen sich die meisten Definitionen und Aussagen auf den Fall einer total geordneten Indexmnege I verallgemeinern, d.h. wenn I eine mit einer Ordnugsrelation $\leq$ ausgestattete Menge ist, so dass für zwei $i, j \in I$ stetst

$i \leq j$ oder $j \leq i$ gilt.) Um die Schreibweise jedoch so einfach wie möglich zu halten, gehen wir von nun an davon aus, dass $I \subseteq \mathbb{N}_0$.

Definition 69. Sei $(\Omega, \mathcal{J})$ ein messbarer Raum, dann heißt eine Familie von Unter-σ-Algebren $(\mathcal{J}_k)_{k\in I\subset\mathbb{N}}$ von $\mathcal{J}$ eine *Filtration* falls $\mathcal{J}_k \subset \mathcal{J}_l$ für alle $k \leq l \in I$. Falls $(\Omega, \mathcal{J})$ mit zudem mit einem Wahrscheinhlichkeitsmaß P ausgestattet ist, nennen wir das Quadrupel $(\Omega, (\mathcal{J}_k)_{k\in I}, \mathcal{J}, P)$ einen *filtrierten* Wahrscheinlichkeitsraum.

Auch hier wollen wir künftig vereinfachend schreiben $(\mathcal{F}_k)_{k\in I} = \mathcal{F}.$, sofern Missverständnisse ausgeschlossen sind. Man beachte also, dass $\mathcal{F}.$ eine Filtration bezeichnet, $\mathcal{F}$ aber für eine feste σ-Algebra steht.

Beispiel 70 (Wiederholter Münzwurf)**.** Zur Beschreibung des dreifachen Münzwurfes führen wir den messbaren Raum $(\Omega, \mathcal{F})$ ein als

$$\Omega = \{(i_1, i_2, i_3) \mid i_j \in \{0, 1\}, j \in \{1, 2, 3\}\}, \quad \mathcal{F} = 2^\Omega.$$

(Falls z.B. die Münze fair ist, könnten wir $(\Omega, \mathcal{F})$ mit der Gleichverteilung als Wahrscheinlicheitsmaß ausstatten, d.h. $P(\{\omega\}) = \frac{1}{8}$ für alle $\omega \in \Omega$.)
Auf Ω können wir die *Koordinatenabbildungen* definieren

$$(X_j)_{j=1,2,3} : \Omega \to \mathbb{R}, \quad X_k\left((\omega_1, \omega_2, \omega_3)\right) := \omega_k.$$

Die Koordinatenabbildung X_k entspricht hier der Auswertung des k-ten Münzwurfs. Für $k = 1, 2, 3$ erhalten wir nun jeweils eine σ-Algebra

$$\mathcal{J}_k := \sigma\left(X_1, ..., X_k\right) = \sigma\left(X_l \mid l \leq k\right).$$

Die σ-Algebra $\mathcal{J}_k$ besteht also aus den nach Beobachtung der ersten k Münzwürfe sinnvoll beschreibbaren Ereignisse und kann wie folgt dargestellt werden.

$$\begin{aligned} \mathcal{J}_1 &= \sigma\left(X_1\right) = \{A_1, A_1^c, \varnothing, \Omega\} \\ \text{mit } A_1 &= \{(0,0,1), (0,1,0), (0,0,0), (0,1,1)\}. \end{aligned}$$

$$\begin{aligned} \mathcal{J}_2 &= \sigma\left(X_1, X_2\right) = \sigma\left(\{A_{11}, A_{12}, A_{21}, A_{22}, \varnothing, \Omega\}\right) \\ \text{mit } A_{11} &= \{(0,0,0), (0,0,1)\} \\ A_{12} &= \{(0,1,0), (0,1,1)\} \\ A_{21} &= \{(1,0,0), (1,0,1)\} \\ A_{22} &= \{(1,1,0), (1,1,1)\} \end{aligned}$$

$$\mathcal{J}_3 = \sigma\left(X_1, X_2, X_3\right) = \mathcal{J} = 2^(\Omega)$$

Dieses Beispiel ist ein Spezialfall der folgenden sehr nützlichen Konstruktion.

Definition 71. Es sei $(X_k)_{k\in I}$ ein auf einem messbaren Raum $(\Omega, \mathcal{F})$ definierter stochastischer Prozess mit Zustandsraum $(E, \mathcal{E})$. Dann heißt $(\mathcal{F}_k^X)_{k\in I}$

$$\mathcal{F}_k^X := \sigma(X_j, j \in I, j \leq k) \subset \mathcal{F}$$

die *vom Prozess X auf $(\Omega, \mathcal{F})$ erzeugte Filtration.*

Diese Definition überträgt sich sinngemäß auf den Fall einer allgemeinen total geordneten Indexmenge I. – Offensichtlich ist dann für jedes $k \in I$ die Zufallsvariable X_k messbar als Abbildung zwischen den messbaren Räumen $(\Omega, \mathcal{F}_k^X) \to (E, \mathcal{E})$.

Definition 72. Es sei $(\Omega, \mathcal{F}., \mathcal{F}, P)$ ein filtrierter Warhscheinlichkeitsraum und $X.$ ein hierauf definierter stochastischer Prozess mit Zustandsraum $(E, \mathcal{E})$. Dann heißt $X.$ an die Filtration $\mathcal{F}.$ *adaptiert*, falls $X_k : \Omega \to E$ messbar bzgl. $\mathcal{F}_k$ ist für alle $k \in I$.

Nach Defintion von $F_\cdot^X$ ist $X.$ dann $\mathcal{F}_\cdot^X$-adaptiert. In der Tat ist $\mathcal{F}_\cdot^X$ die kleinste Filtration mit dieser Eigenschaft.

Zur Erläüterung der Adaptiertheitsbedingung stellen wir hier das sogenannte *Faktoriserungslemma* vor, das man leicht auf elementare Weise verifizieren kann (einen abstrakten Beweis findet man als Anwendung des Monotone Klassen-Argumentes im Anhang).

Satz 73 (Faktoriserungslemma)**.** *Es sei $X : \Omega \mapsto E$ eine messbare Abbilung zwischen den messbaren Räumen $(\Omega, \mathcal{F})$ und $(E, \mathcal{E})$. Falls $Z : \Omega \to \mathbb{R}$ messbar ist bzgl. der von X erzeugten σ-Algebra $\sigma(X)$, so gibt es eine messbare Abbildung $h : E \to \mathbb{R}$, so dass $Z = h(X)$.*

$$\begin{tikzcd} (\Omega, \mathcal{J}) \arrow[r, "X"] \arrow[d, "Z"'] & (E, \mathcal{E}) \arrow[dl, "h"] \\ (\mathbb{R}, \mathcal{B}(\mathbb{R})) & \end{tikzcd}$$

Wenn wir also etwa in der Situation von Beispiel 70 verlangen, dass eine Zufallsvariable $\mathcal{J}_2$-messbar sei, ist dies gleichbedeutend, dass sie als Funktion der ersten beiden Münzwürfe dargestellt werden kann. Entsprechend gibt die von einem Prozess $X.$ erzeugte Filtration $\mathcal{F}_\cdot^X$ die wachsenden Mengen der durch fortschreitende Beobachtung von $X.$ beschreibbaren Ereignisse an.

Definition 74. Ein stochastischer Prozess $M. = (M_k)_{k \in I}$ mit $M_k : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ defniert auf einem filtrierten Wahrscheinlichkeitsraum $(\Omega, (\mathcal{J}_k)_{k \in I}, \mathcal{J}, P)$ heißt *Martingal*, falls

(1) M_k integrierbar, d.h. $\min \{\mathbb{E}(M_+), \mathbb{E}(X_-)\} < \infty$
(2) M_k messbar bzgl. $\mathcal{J}_k$
(3) $\mathbb{E}_P(M_k \mid \mathcal{J}_k) = M_k$ P-f.s. falls $k \geq l$

Bemerkung 75. Genauer sagen wir, dass $M. = (M_k)$ ein $(\mathcal{J}., P)$-Martingal ist, wobei $\mathcal{J}. = (\mathcal{J}_k)_{k \in I}$ die zugrunde gelegte Filtrierung bezeichnet. – Wenn keine Gefahr einer Verwechselung besteht, verzichten wir jedoch häufig auf die explizite Nennung von $\mathcal{J}.$ und P.

Definition 76. Sei $(S_k)_{k \in I} = S.$ heißt $(\mathcal{J}., P)$-*Submartingal*, falls

(1) $\mathbb{E}\left(S_k^+\right) < \infty$
(2) S_k messbar bzgl. $\mathcal{J}_k$
(3) $\mathbb{E}_p(S_l \mid \mathcal{J}_k) \geq S_k$ $P-$f.s. für alle $l \geq k$.

$(S.)$ heißt $(\mathcal{J}., P)$-*Supermartingal*, falls $(-S_k)_{k \in I}$ ein Submartingal ist.

Beispiel 77. Gegeben seien $(\Omega, \mathcal{J}, P), (X_k)_{k \in \mathbb{N}}$ und eine Folge von unabhängigen Zufallsvariablen mit $\mathbb{E}(X_k) = 0$. Wir setzen

$$M_k := \sum_{k=1}^{l} X_k, \mathcal{J}_k := \sigma(M_1, ..., M_k) = \mathcal{J}_k^{M.}$$

Dann ist M ein $\left(\mathcal{J}^X., P\right)$-Martingal, denn

$$\begin{aligned}
\mathbb{E}\left(M_k \mid \mathcal{J}_k\right) &= \mathbb{E}\left(\sum_{j=1}^{l} X_j + \sum_{j=l+1}^{k} X_j \mid \sigma\left(X_1, ..., X_l\right)\right) \\
&= \mathbb{E}\left(\underbrace{\sum_{j=1}^{l} X_j}_{=M_l} \mid \underbrace{\sigma\left(X_1, ..., X_l\right)}_{=\mathcal{J}_l}\right) + \mathbb{E}\left(\underbrace{\sum_{j=l+1}^{k} X_j \mid \sigma\left(X_1, ..., X_l\right)}_{\text{zueinander unabh.}}\right) \\
&= M_l + \mathbb{E}\left(\sum_{j=l+1}^{k} X_j\right) \\
&= M_l + \sum_{j=l+1}^{k} \mathbb{E}\left(X_j\right) \\
&= M_l
\end{aligned}$$

Analog falls $\mathbb{E}\left(X_k\right) \geq 0$ für alle k. Dann gilt

$$\mathbb{E}\left(M_k \mid \mathcal{J}_l\right) \geq M_l$$

folglich ist $M.$ dann ein $\left(\mathcal{J}^X., P\right)$-Submartingal.

Beispiel 78 (Verdoppelungsstrategie)**.** Wir betrachten den wiederholten unabhängigen Wurf einer faieren Münze, modelliert als eine i.i.d. Folge von fairen Benulli-Variablen $\lambda_n \in \{1, -1\}$. Zwei Spieler A und B vereinbaren jeweils vor dem n-ten Wurf einen Wetteinsatz von ξ_n Euro, zu zahlen vom Verlierer des n-ten Münzwurfs an den Gewinner. Der Gesamtgewinn bzw. -verlust nach n Runden für Spieler A sind damit durch den Ausdruck

$$Z_n := \sum_{k=1}^{n} \lambda_k \cdot \xi_n, \quad n \geq 1.$$

beschrieben, mit der Vereinbarung, dass $Z_0 := 0$. Die *Verdoppelungsstrategie* besteht nun darin, dass Spieler A als Wetteinsatz für die n-te Runde den Betrag

$$\xi_n := 2 \max(-Z_{n-1}, 0)$$

wählt, d.h. er setzt in der $n{+}1$-ten Runde das doppelte seiner bisherigen Verluste, um diese im Gewinnfalle auszugleichen und mit einem Reingewinn das Spiel zu beenden. Es sei $\mathcal{J}_n := \sigma\left(\lambda_1, ..., \lambda_n\right)$ die von den Realisierungen der Münzwürfe erzeugte Filtrierung, dann ist offenbar $(Z.)$ an $(\mathcal{J}.)$ $-$adaptiert. Der Prozess $\xi.$ ist $\mathcal{J}.$-vorhersagbar. Wir behaupten, dass $Z.$ ein $\mathcal{J}.$-Martingal ist, denn

$$\begin{aligned}
\mathbb{E}\left(Z_{n+1} \mid \mathcal{J}_n\right) &= \mathbb{E}\left(Z_n + \xi_{n+1}\lambda_{n+1} \mid \sigma\left(\lambda_1, ..., \lambda_n\right)\right) \\
&= \mathbb{E}\left(\xi_{n+1}\lambda_{n+1} \mid \sigma\left(\lambda_1, ..., \lambda_n\right)\right) + \underbrace{\mathbb{E}\left(Z_n \mid \sigma\left(\lambda_1, ..., \lambda_n\right)\right)}_{=Z_n} \\
&= \mathbb{E}\left(\lambda_{n+1}\xi_{n+1} \mid \sigma\left(\lambda_1, ..., \lambda_n\right)\right) + Z_n \\
&= \xi_{n+1}\mathbb{E}\left(\underbrace{\lambda_{n+1} \mid \sigma\left(...\right)}_{\text{unabh.}}\right) + Z_n \\
&= \xi_{n+1}\mathbb{E}\left(\lambda_{n+1}\right) + Z_n \\
&= \xi_{n+1}0 + Z_n = Z_n.
\end{aligned}$$

Bemerkung 79. In der Tat ist für jede Vorhersagbare Strategie $\xi_\cdot$, d.h. $\xi_{n+1} = F_{n+1}(\lambda_1, \cdots, \lambda_n)$ der resultierende Gesamtgewinnprozess $Z_\cdot$ ein Martingal.

3.2. **Beispiel: Martingale bei Markovketten.**

Definition 80. Sei P eine Übergangsmatrix auf E einem abszählbaren Zustandsraum, d.h. $P = (P)_{(u,v)\in E\times E}$. Dann heißt $f : E \to \mathbb{R}$ *P-harmonisch*, falls

$$f(x) = \sum_{y\in E} f(y)P(x,y)$$

Bemerkung 81. Im Kontinuum entspricht dies der Mittelwerteigenschaft für harmonische Funktionen im euklidischen Raum. Hierbei heißt eine reelle Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ harmonisch, wenn

$$\Delta f := \partial_x^2 f + \partial_y^2 f = 0.$$

Harmonische Funktionen haben die Mittelwerteigenschaft (vergl. auch Cauchy'scher Integalsatz aus der komplexen Analysis.)

Satz 82. *Sei $f : E \to \mathbb{R}$ f beschränkt. Dann ist f P-harmonisch genau dann, wenn $(M_k) = f\left(X_k\right)$ ein $(\mathcal{F}_\cdot^X, P_\lambda)$-Martingal ist für jede Startverteilung λ.*
Lemma 83.

$$\mathbb{E}_\lambda\left(f\left(X_{t+1}\right) \mid \sigma\left(X_0, ..., X_t\right)\right) = \mathbb{E}_{X_t}\left(f\left(X_1\right)\right) \quad P_\lambda\text{-f.s.}$$

für alle $f : E \to \mathbb{R}$.

Beweis. In Kombination mit einem Monotone Klassen-Argument(siehe Anhang) reicht es aus, den Fall $f = 1_A$ mit $A \subset E$ und hier sogar nur den Spezialfall, dass $A = \{y\}, y \in E$, zu betrachten. Mit der elementaren Markov-Eigenschaft gilt zunächst einmal, dass

$$\mathbb{E}_\lambda\left(1_{\{y\}}\left(X_{t+1}\right) \mid \sigma\left(X_0, ..., X_t\right)\right) = \mathbb{E}_\lambda\left(1_{\{y\}}\left(X_{t+1}\right) \mid \sigma\left(X_t\right)\right)$$

Sei nun $B \subset \sigma\left(X_t\right)$, also wieder mit monoton Klassen o.B.d.A.

$$B = \left\{X_t \in C\right\}, C \subset E$$

Dann gilt

$$
\begin{aligned}
\mathbb{E}_\lambda\left(1_{\{y\}}\left(X_{t+1}\right) 1_C\left(X_t\right)\right) &= \sum_{c\in C} P_\lambda\left(X_{t+1}=y; X_t=c\right)\\
&= \sum_{c\in C} P_\lambda\left(X_{t+1}=y \mid X_t=c\right) P_\lambda\left(X_t=c\right)\\
&= \sum_{c\in C} P_\lambda\left(X_t=c\right) P(c,y)\\
&= \sum_{c\in C} \mathbb{E}_\lambda\left(1_{\{c\}}\left(X_t\right) P(c,y)\right)\\
&= \sum_{c\in C} \mathbb{E}_\lambda\left(1_{\{c\}}\left(X_t\right) P_{X_t}\left(X_1=y\right)\right)\\
&= \mathbb{E}_\lambda\left(1_C\left(X_t\right) P_{X_t}\left(X_1=y\right)\right)\\
&= \mathbb{E}_\lambda\left(\mathbb{E}_{X_t}\left(1_{\{y\}}\left(X_1\right)\right) 1_C\right),
\end{aligned}
$$

was zeigt, dass $E_\lambda(1_{\{y\}}(X_{t+1})|\sigma(X_t)) = E_{X_t}(1_{\{y\}}(X_1))$ fast sicher. □

Beweis von Satz 82. Sei f harmonisch, $M_k := f\left(X_k\right)$, λ bel. Startverteilung:

$$
\begin{aligned}
\mathbb{E}_\lambda\left(M_{k+1} \mid \mathcal{J}_k^X\right) &= \mathbb{E}_\lambda\left(f\left(X_{k+1}\right) \mid \sigma\left(X_0,\dots,X_k\right)\right)\\
&\overset{\text{Lemma 39}}{=} \mathbb{E}_{X_k}\left(f\left(X_1\right)\right)\\
&= \sum_{y\in E} f(y) p_{X_k,y}\\
&\overset{f\text{ harmonisch}}{=} f\left(X_k\right)\\
&= M_k
\end{aligned}
$$

Wähle umgekehrt beispielsweise $\lambda = \delta_{X_0}$, d.h. deterministischer Start in X_0.

$$
\underbrace{\mathbb{E}_{X_0}\left(\underbrace{f\left(X_1\right)}_{M_1} \mid \sigma\left(X_0\right)\right)}_{=\sum f(y)P(x_0,y)} = f\left(X_0\right) = M_0
$$

also ist f harmonisch. □

Bemerkung 84. Eine analoge Aussage gilt für harmonische Funktionen und die Brown'sche Bewegung als Prozess im kontinuierlichen Zustandsraum $\mathbb{R}^d$, siehe Abschnitt 6.

3.3. **Elementare Eigenschaften von Martingalen.**

Satz 85.

(1) Falls $X.$ ein $\mathcal{J}.$-Martingal ist und an $\mathcal{G}.$ adaptiert für eine Filtration $\mathcal{G}.$ mit $\mathcal{G}_t \subset \mathcal{G}_t$ für alle t, so ist X ein $\mathcal{G}.$-Martingal.
(2) M, N seien $\mathcal{J}.$-Martingale. Dann ist auch $aM + bN$ $\mathcal{J}.$-Martingal
(3) X, Y seien Submartingale. Dann ist

$$Z_t(\omega) = X_t(\omega) \vee Y_t(\omega) \qquad \text{(Supremum)}$$

Submartingal bzgl. $\mathcal{J}.$

(4) Falls $X.$ Submartingal, φ konvex und monoton wachsend, dann ist $Y_t := \varphi(X_t)$ Submartingal
(5) Falls $X.$ Martingal und ϕ konvex, dann ist $Y_t := \phi(X_t)$ Submartingal

Beweis. Die Eigenschaft 1 folgt aus der Turmeigenschaft der bedingten Erwartung, denn

$$\begin{aligned}\mathbb{E}_p\left(X_{t+s} \mid \mathcal{G}_t\right) &\overset{\mathcal{G}_t \subseteq \mathcal{J}_t}{=} \mathbb{E}_p\left(\underbrace{\mathbb{E}_p\left(X_{t+s} \mid \mathcal{J}_t\right)}_{=X_t} \mid \mathcal{G}_t\right)\\ &= \mathbb{E}\left(X_t \mid \mathcal{G}_t\right)\\ &= X_t.\end{aligned}$$

Die Aussagen 4 unf 5 folgen aus der Jensen'schen Ungleichung für die bedingte Erwartung. Zum Beweis von 3 nutzen wir die triviale Aussage

$$X(\omega) \leq X(\omega) \vee Y(\omega)$$

zusammen mit der Submartingaleigenschaft von X bzw. Y sowie der Monotonie der bedingten Erwartung und erhalten

$$X_t \leq \mathbb{E}\left(X_{t+s} \mid \mathcal{J}_t\right) \leq \mathbb{E}\left(X_{t+s} \vee Y_{t+s} \mid \mathcal{J}_t\right) \qquad P\text{-f.s.}$$

und

$$Y_t \leq \mathbb{E}\left(Y_{t+s} \mid \mathcal{J}_t\right) \leq \mathbb{E}\left(X_{t+s} \vee Y_{t+s} \mid \mathcal{J}_t\right) \qquad P\text{-f.s.}$$

Somit gilt dann auch

$$X_t(\omega) \vee Y_t(\omega) \leq \mathbb{E}\left(X_{t+s} \vee Y_{t+s} \mid \mathcal{J}_t\right) \qquad P\text{-f.s.}$$

□

Ein sehr nützliches Hilfsmittel ist die Möglichkeit, ein Submartingal in ein Martingal und einen monoton wachsenden Prozess zu zerlegen. Hierbei nennen wir einen Prozess $A.$ monoton wachsend (fast sicher), falls $A_n \geq A_{n-1}$ fast sicher. Zudem benötigen wir für die Eindeutigkeit der Zerlegung noch den folgenden Begriff.

Definition 86. Gegeben sei ein filtrierter Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}_., \mathcal{F}, P)$. Dann heißt ein auf $(\Omega, \mathcal{F}, P)$ definierter Prozess *vorhersagbar*, falls X_t messbar ist bezgl. $\mathcal{F}_{t-1}$ für alle $t \in \mathbb{N}$.

Falls also z.B. $\mathcal{F}_. = \mathcal{F}_.^Z$ die von einem Prozess $Z.$ erzeugte Filtration ist, so bedeutet die $\mathcal{F}_.^Z$-Vorhersagbarkeit von $X.$ gerade, dass X_t bereits durch die Werte von $Z_1, \cdots, Z_{t-1}$ festgelegt ist.

Satz 87 (Doob-Zerlegung)**.** *Sei $(X_n)_{n\in\mathbb{N}}$ ein Submartingal mit $\mathbb{E}(|X_n|) < \infty$. Dann existieren ein vorhersagbarer wachsender Prozess $(A_n)_{n\in\mathbb{N}}$, $A_0 = 0$, und ein Martingal $(M_n)_{n\in\mathbb{N}}$, so dass $X_n = M_n + A_n$. Diese Zerlegung ist eindeutig.*

Beweis. Sei $S_i := \mathbb{E}\left(X_i \mid \mathcal{J}_{l-1}\right) - X_{i-1} \geq 0$ wegen der Submartingaleigenschaft, $i \geq 1$. S_i ist $\mathcal{J}_{i-1}-$messbar. Definiere weiter $A_i := \sum_{k=1}^{i} S_k$, $i \geq 1, A_0 = 0$ und $M_i := X_i - A_i, i \geq 1, M_0 = X_0$. Dann ist:

$$\begin{aligned}M_{i+1} - M_i &= X_{i+1} - X_i - (A_{i+1} - A_i)\\ &= X_{i+1} - X_i - \left(\mathbb{E}\left(X_{i+1} \mid \mathcal{J}_i\right) - X_i\right)\\ &= X_{i+1} - \mathbb{E}\left(X_{i+1} \mid \mathcal{J}_i\right)\end{aligned}$$

Folglich

$$\mathbb{E}\left(M_{i+1}-M_i \mid \mathcal{J}_i\right) \;=\; \mathbb{E}\left(X_{i+1} \mid \mathcal{J}_i\right)-\mathbb{E}\left(X_{i+1} \mid \mathcal{J}_i\right)=0.$$

Mithin ist $(M_k)_{k\in\mathbb{N}_0}$ Martingal, denn aus $\mathbb{E}\left(M_{i+1} \mid \mathcal{J}_i\right)=M_i$ folgt auch $\mathbb{E}\left(M_{k+i} \mid \mathcal{J}_i\right)=M_l$ per Induktion. Also ist $X_i:=M_i+A_i$ eine Zerlegung wie gewünscht.
Für die Eindeutigkeit schreiben wir $X_i-X_{i-1}=M_i-M_{i-1}+A_i-A_{i-1}$, d.h.

$$\begin{aligned}
A_i-A_{i-1} &= \mathbb{E}\left(A_i \mid \mathcal{J}_{i-1}\right)-A_{i-1}\\
&= \mathbb{E}\left(A_i-A_{i-1} \mid \mathcal{J}_{i-1}\right)\\
&= \mathbb{E}\left(X_i-X_{i-1} \mid \mathcal{J}_{i-1}\right)-\mathbb{E}\left(\underbrace{M_i-M_{i-1} \mid \mathcal{J}_{i-1}}_{=0}\right)\\
&= \mathbb{E}\left(X_i \mid \mathcal{J}_{i-1}\right)-X_{i-1}
\end{aligned}$$

Mithin ist A_i-A_{i-1} eindeutig bestimmt durch $X.$, und somit auch auch

$$A_l=A_0+\sum_{j=1}^{l}\left(A_j-A_{j-1}\right)$$

durch Festlegun von $A_0:=0$. Schließlich ist damit auch $M.$ durch $M_k:=X_k-A_k$ eindeutig bestimmt. □

3.4. **Optional Sampling und Optional Stopping.** Es stellt sich heraus, dass sich die (Sub-)Martingaleigenschaft $E(X_t|\mathcal{F}_s)\geq X_s$ von deterministischen Zeitpunkten $s\leq t$ auf zufällige gewählte Zeitpunkte fortsetzen lässt, d.h. wenn s und t ersetzt werden durch Zufallsvariablen $S\leq T:\Omega\to\mathbb{N}$, welche als *Zufallszeiten* interpretiert werden. Wichtig dabei ist, dass die verwendeten Zufallszeiten nicht in die Zukunft schauen können, was durch den Begriff der *Stoppzeit* wie folgt formalisiert wird.

Definition 88. Sei $\left(\Omega,\mathcal{J},(\mathcal{J}_k)_{k\in I}\right)$ ein filtrierter W-Raum. Dann heißt $\tau:\Omega\to I$ eine $(\mathcal{J}.)-$*Stoppzeit*, falls

$$\{\tau\leq t\}\in\mathcal{J}_t, \forall t\in I$$

Man kann $\tau:\Omega\to\mathbb{R}$, $\tau(\omega)\in I$ als mathematische Beschreibung eines Alarmmechanismus interpretieren. Das Ereinis $\{\tau\leq t\}$ entsteht aus sämtlichen Realisierungen des Zufallsparameters ω, bei denen Alarm spätestens zum Zeitpunkt t ausgelöst wurde. Falls z.B. $\mathcal{J}=\mathcal{F}^X$, so besagt die obige Forderung, dass allein durch Beobachtung von $X.$ bis einschl. zum Zeitpunkt t entscheidbar ist, ob bis hierhin der Alarm auszulösen war. Insofern beschreibt eine Stoppzeit τ einen Alarmmechanismus, der nicht in die Zukunft schauen kann.

Beispiel 89. (1) Das typische Beispiel einer Stoppzeit ist die *Trefferzeit einer Menge* durch einen adaptierten Prozess. Sei $X.$ eine reeller $(\mathcal{J}.)$-adaptierter Prozess und $A\in\mathcal{B}(\mathbb{R})$. Dann definieren wir

$$\tau:\Omega\to\mathbb{R},\quad \tau(\omega):=\inf\left\{l\in I \mid X_l\in A\right\}.$$

Dann ist τ ist eine $(\mathcal{J})$-Stoppzeit, denn

$$\begin{aligned}\{\tau \le t\} &= \bigcup_{j=0}^{t} \{\tau = j\} \\ &= \bigcup_{j=0}^{t} \{\tau \le j\} \\ &= \bigcup_{j=0}^{t} \underbrace{X_j^{-1}(A)}_{\in \mathcal{J}_j \subset \mathcal{J}_k} \in \mathcal{J}_t.\end{aligned}$$

(2) Als einprägsames Beispiel einer Zufallszeit, die keine Stoppzeit ist, betrachte man in der Situation wie oben die *letzte Besuchszeit* einer Menge

$$\sigma : \Omega \to \mathbb{R}, \quad \sigma(\omega) = \sup \{j \in I \mid X_j \in A\}.$$

Offensichttlich ist dann $\{\sigma \le t\} \notin \mathcal{J}_t$, d.h. man muss in die Zukunft jenseits von t schauen, um zu sehen, ob $\tau \le t$ vorliegt oder nicht.

Satz 90. *Falls S, T zwei $(\mathcal{J}.)$-Stoppzeiten sind, so auch $S \wedge T$ (Infimum) und $S \vee T$ (Supremum), bzw. für $S, T \ge 0$ ebenfalls $S + T$.*

Beweis. Beispielsweise gilt $\{S \vee T \le t\} = \{S \le t\} \wedge \{T \le t\} \in \mathcal{J}_t \wedge \mathcal{J}_t$. □

Definition 91. Für eine $(\mathcal{J}.)$-Stoppzeit τ heißt

$$\mathcal{J}_\tau := \{A \in \mathcal{J} \mid \{\tau \le t\} \cap A \in \mathcal{J}_t\}$$

die *σ-Algebra der τ-Vergangenheit.*
Satz 92.

(1) $\mathcal{J}_\tau$ ist eine $\sigma-$Algebra in $\mathcal{J}$.
(2) $S \le T \Longrightarrow \mathcal{J}_S \subseteq \mathcal{J}_T$
(3) $\mathcal{J}_{S \wedge T} = \mathcal{J}_S \wedge \mathcal{J}_T$

Beweis. 1) $A \in \mathcal{J}_\tau$. Zeige $A^c \in \mathcal{J}_\tau$. Sei also $A \in \mathcal{J}$. Dann ist auch $A^c \in \mathcal{J}$.

$$A^c \cap \{\tau \le t\} = \underbrace{\{\tau \le t\}}_{\in \mathcal{J}_t} \setminus \underbrace{(A \cap \{\tau \le t\})}_{\in \mathcal{J}_t}$$

2) $\{S \le T\} \Longrightarrow \{T \le t\} \subset \{S \le t\}$, $A \in \mathcal{J}_S$:

$$A \cap \{T \le t\} = \underbrace{A \cap \{S \le t\}}_{\in \mathcal{J}_t} \cap \underbrace{\{T \le t\}}_{\in \mathcal{J}_t}$$

3) folgt aus 2). □

Lemma 93. *Für einen $(\mathcal{J}_i)_{i \in I}$-adaptierten Prozess $(X_i)_{i \in I}$ mit Zustandsraum E sowie eine $\mathcal{J}.$-Stoppzeit $\tau : \Omega \to I$ ist*

$$X_\tau : \Omega \to E, \quad X_\tau(\omega) := X_{\tau(\omega)}(\omega)$$

eine $\mathcal{F}_\tau$-messbare Zufallsvariable.

Beweis. In der Tat, für $A \in \mathcal{E}$ und $i \in I$ gilt

$$\begin{aligned} X_\tau^{-1}(A) \cap \{\tau \le i\} &= \bigcup_{k \in I, k \le i} X_\tau^{-1}(A) \cap \{\tau = i\} \\ &= \bigcup_{k \in I, k \le i} \underbrace{X_k^{-1}(A) \cap \{\tau = k\}}_{\in \mathcal{J}_k \subset \mathcal{J}_i} \in \mathcal{J}_i. \end{aligned}$$

□

Satz 94. *(Optimal Sampling Theorem) Falls $(X_n)_{n \in \mathbb{N}_0}$ ein $(\mathcal{J}_n)_{n \in \mathbb{N}_0}$ ein Submartingal ist und*

$$0 \le S \le T \le n_0$$

zwei beschränkte Stoppzeiten sind, dann gilt

$$\mathbb{E}\left(X_T \mid \mathcal{J}_S\right) \ge X_S \text{ fast sicher.}$$

Analog mit "$\le$" für für Supermartingale bzw. mit "$=$" für Martingale.

Beweis. Wir geben den Beweis in drei Schritten.

Schritt 1: Sei X Martingal, $T = t$ deterministisch, $T \ge S$. Dann ist zu zeigen, dass

$$\mathbb{E}\left(X_T \mid \mathcal{J}_S\right) = X_S.$$

Zu diesem Zweck zeigen wir, dass $\mathbb{E}\left(X_T \cdot 1_A\right) = \mathbb{E}\left(X_S \cdot 1_A\right)$ für alle $A \in \mathcal{J}_S$ wie folgt

$$\begin{aligned} \mathbb{E}\left(X_S \cdot 1_A\right) &= \sum_{k=0}^{T} \mathbb{E}\left(X_T \cdot 1_{S=k} 1_A\right) \\ &= \sum_{k=0}^{T} \mathbb{E}\left(X_T 1_{S=k} 1_A\right) \\ &= \sum_{k=0}^{T} \mathbb{E}\left(X_T 1_{\underbrace{\{S=k\} \cap A}_{\in \mathcal{J}_k}}\right) \\ &\overset{\text{Martingaleigenschaft}}{=} \sum_{k=0}^{T} \mathbb{E}\left(X_k 1_{\{S=k\} \cap A}\right) \\ &= \sum_{k=0}^{T} \mathbb{E}\left(X_S 1_{\{S=k\} \cap A}\right) \\ &= \mathbb{E}\left(X_S 1_{A \cap \{S \le T\}}\right) \\ &= \mathbb{E}\left(X_S 1_A\right). \end{aligned}$$

Schritt 2: Sei $X.$ ein Martingal, $S \le T \le n_0$. Dann folgt gemäß Schritt 1 und der Turmeigenschaft der bedingten Erwartung, dass

$$\begin{aligned} E(X_T | \mathcal{J}_S) &= E(E(X_{n_0} | \mathcal{J}_T) | \mathcal{J}_S) \\ &= E(X_{n_0} | \mathcal{J}_S) = X_S \text{ f.s.} \end{aligned}$$

Schritt 3: Sei $X.$ ein Sub-Martingal. Es sei $X. = M. + A.$ die zugehörige Doob-Zerlegung, dann schließen wir unter Verwendung von Schritt 2, der Monotonie von $A.$ und der Monotone der bedingten Erwartung, dass

$$\begin{aligned} X_S &= M_S + A_S = E(M_T|\mathcal{J}_S) + A_S \\ &= E(M_T + A_S|\mathcal{J}_S) \leq E(M_T + A_T|\mathcal{F}_S). \end{aligned}$$

□

Bemerkung 95. Die Bedingung, dass T beschränkt ist, kann i.A. nicht weggelassen werden. Hierzu betrachte man erneut das Beispiel der Verdoppelungsstrategie (Beispiel 78). Führen wir die Zeit des ersten Gewinns

$$T := \inf\{k \in \mathbb{N} \mid Z_k > 0\}$$

ein, so macht man sich schnell klar, dass zwar $T < \infty$ f.s. Doch T ist nicht beschränkt, d.h. es existiert kein $C > 0$, s.d.

$$T(\omega) < C, \forall \omega,$$

denn andernfalls wäre $0 = X_0 = E(X_T|\mathcal{J}_0) = E(X_T)$, im Widerspruch zu $X_T > 0$ fast sicher.

Das Optional Sampling Theorem handelt vom Vergleich des Prozesses an zwei durch gewählte Stoppzeiten S und T festgelegten Zeitpunkten (engl. Sample = Auswählen). Dem gegenüber steht das verwandte sogenannte Optional Stopping Theorem, das die Stabilität der Sub- bzw. Supermartinaleigenschaft von Prozessen unter der *Stopp-Transformation* der Pfade behauptet.

Definition 96 (Stoppen eines Prozesses)**.** Für eine Zufallszeit $\tau : \Omega \to I$ und einen Stochastischen Prozess $(X_i)_{i\in I}$ heißt der Prozess $(X_i^\tau)_{i\in I}$ mit

$$X_k^T(\omega) := X(\omega)_{k\wedge\tau(\omega)} = \begin{cases} X_k(\omega) & \text{, falls } T(\omega) \geq k \\ X_{T(\omega)}(\omega) & \text{, sonst} \end{cases}$$

die *τ-Stoppung von X^τ* bzw. einfach der *an τ gestoppte Prozess.*

Satz 97 (Optional Stopping)**.** *Falls $(X_k)_{k\in\mathbb{N}}$ ein $\mathcal{J}.$-adaptiertes Submartingal ist und $T \leq n$ eine beschränkte Stoppzeit, so ist $X_\cdot^T$ wieder ein $\mathcal{J}.$-Submartingal.*

Zum Beweis benutzen wir die folgende Charakterisierung von Submartingalen.

Lemma 98. *Sei $(X_k)_{k\in\mathbb{N}}$ ein stoch. Prozess, $\mathcal{J}.$-adaptiert, $\mathbb{E}(X_+) \wedge \mathbb{E}(X_-) < \infty$. Dann sind äquivalent:*

(1) $X.$ ist ein $\mathcal{J}.$-Submartingal.
(2) $\forall S \leq T$ beschränkte Stoppzeiten gilt

$$\mathbb{E}(X_S) \leq \mathbb{E}(X_T).$$

Beweis. "$\Rightarrow$": Mit Optimal Sampling folgt

$$\mathbb{E}(X_T \mid \mathcal{J}_S) \geq X_S \qquad \text{f.s.},$$

woraus durch Übergang zum Erwartungswert (und der Turmeigenschaft der bedingten Erwartung) folgt, dass

$$E(X_T) = \mathbb{E}(\mathbb{E}(X_T \mid \mathcal{J}_S)) \geq \mathbb{E}(X_S)$$

"$\Leftarrow$" $0 \leq u < v$ seien zwei deterministische Zeitpunkte, $A \in \mathcal{J}_v$, definiere $T : \Omega \to \mathbb{R}$ mittels $T(\omega) = u1_A + v1_{A^c}$. Dann ist

$$\{T \leq w\} = \begin{cases} \varnothing & \text{, falls } w < u \\ A & \text{, falls } u \leq w < v \\ \Omega & \text{, falls } w \geq v \end{cases}$$

$$\Longrightarrow \{T \leq w\} \in \mathcal{J}_w$$

Somit ist T eine $\mathcal{J}_.$-Stoppzeit mit $T \leq v < \infty$ beschränkt. Folglich gilt laut Voraussetzung

$$\begin{aligned} \Longrightarrow \mathbb{E}(X_v) &\geq \mathbb{E}(X_T) \\ &= \int_A X_T dP + \int_{A^c} X_T dP \\ &= \int_A X_u dP + \int_{A^c} X_v dP, \end{aligned}$$

also

$$\int_A X_v dP \geq \int_A X_v dP \quad \forall A \in \mathcal{J}_u.$$

Dies ist aber gerade die Submartingaleigenschaft von $X_.$. □

Beweis vom Satz 97. Seien $\sigma \leq \tau \leq c$ zwei beschränkte Stoppzeiten. Dann ist

$$X_\sigma^T = X_{T\wedge\sigma}, X_\tau^T = X_{T\wedge\tau}, T\wedge\sigma \leq T\wedge\tau \leq c$$

und $T\wedge\sigma, T\wedge\tau$ sind $\mathcal{J}_.$-Stoppzeiten. Somit gilt

$$\begin{aligned} \mathbb{E}\left(X_\tau^T\right) &= \mathbb{E}\left(X_{T\wedge\tau}\right) \\ &\overset{\text{optional Sampling}}{\geq} \mathbb{E}\left(X_{T\wedge\sigma}\right) \\ &= \mathbb{E}\left(X_\sigma^T\right) \end{aligned}$$

Insgesamt ist der Prozess $X_.^T$ also $\mathcal{J}_.$-adaptiert und es gilt $\mathbb{E}\left(X_\tau^T\right) \geq \mathbb{E}\left(X_\sigma^T\right)$ für alle $\sigma \leq \tau$ beschränkte Stoppzeiten. Nach dem vorausgehenden Lemma ist $X_.^T$ also ein $\mathcal{J}_.$-Submartingal. □

Bemerkung 99. $X_.^T$ ist auch ein $\mathcal{J}_.^T$-Submartingal, wobei

$$\mathcal{J}_n^T := \mathcal{J}_{T\wedge n} \subset \mathcal{J}_n,$$

denn $X_.^T$ ist ein $\mathcal{J}_.$-Martingal, und bereits $\mathcal{J}_.^T$-adaptiert.

3.5. **Anwendung: Optimales Stopproblem und Snell'sche Einhüllende.** Man stelle sich ein Spiel vor, bei dem ein Spieler wiederholt einen Zufallsmechanismus betätigen kann und von Runde zu Runde entscheiden darf, das Spiel mit dem aktuellen Wurf als Gewinn zu beenden oder stattdessen erneut den Mechanismus zu betätigen.

Eine mathematische Formulierung des Problems sieht dann wie folgt aus. Sei $X_0, X_1, ..., X_N$ ein $\mathcal{J}_.$-adaptierter Prozess. Wir suchen eine optimale $\mathcal{J}_.$-Stoppzeit $\tau \leq N$, so dass

$$\mathbb{E}(X_\tau) \to \max.$$

Diese Aufgabe wird als *Problem des optimalen Stoppens* bezeichnet.

Definition 100. Es sei $(X_k)_{k=0,\cdots,N}$ dann heißt der rekursiv definierte Prozess $(Z_k)_{k=0,\cdots,N}$

$$\begin{aligned} Z_N &:= X_N, \\ Z_{n-1} &:= X_{n-1} \vee \mathbb{E}\left(Z_n \mid \mathcal{J}_{n-1}\right), \end{aligned}$$

die *Snell'sche Einhüllende* von $X.$ (zur Filtration $\mathcal{J}.$ und Zeithorizont N).

Satz 101. *$(Z_k)_{k=0,\dots,N}$ ist das minimale Supermartingal, das $X.$ dominiert.*

Beweis. $Z_n := \max\left(\mathbb{E}\left(Z_{n+1} \mid \mathcal{J}_n\right), X_n\right) \geq \mathbb{E}\left(Z_{n+1} \mid \mathcal{J}_n\right)$, also ist $Z.$ ein Supermartinal.

Sei $\tilde{Z}.$ Supermartingal mit $\tilde{Z}_n \geq X_n$ f.s. für alle $n = 0, ..., N$. Wir beweisen durch Rückwartsinduktion, dass dann auch $\tilde{Z}_k \geq Z_k$ für alle $k = N, N_1, \cdots, 0$ gilt. Für $k = N$ ist das offensichtlich richtig, denn

$$\tilde{Z}_N \geq X_N = Z_N.$$

Falls wir bereits gezeigt haben, dass

$$\tilde{Z}_N \geq Z_N, ..., \tilde{Z}_M \geq Z_M$$

folgt unter der Verwendung der Submartingaleigenschaft von $\tilde{Z}.$ under der Monotonie der bedingten Erwartung hieraus, dass

$$\tilde{Z}_{M-1} \geq \mathbb{E}\left(\tilde{Z}_M \mid \mathcal{J}_{M-1}\right) \geq \mathbb{E}\left(Z_M \mid \mathcal{J}_{M-1}\right).$$

Nach Voraussetzung gilt ferner

$$\tilde{Z}_{M-1} \geq X_{M-1}$$

so dass zusammengenommen auch

$$\tilde{Z}_{M-1} \geq X_{M-1} \vee \mathbb{E}\left(Z_M \mid \mathcal{J}_{M-1}\right).$$

Die rechte Seite stimmt aber nach Konstruktion mit Z_{M-1} überein. $\square$

Lemma 102. *Fur $\tau := \inf\{k \mid Z_k = X_k\}$ ist $Z.^\tau$ ist ein Martingal.*

Beweis. Aus der $\mathcal{J}.$-Adaptiertheit der Prozesse $Z.$ und $X.$ folgt, dass τ eine $\mathcal{J}.$-Stoppzeit ist. Zudem gilt wegen $Z_N = X_N$ offenbar, dass $\tau \leq N$. Für die Inkremente von $Z_.^\tau$ gilt

$$\begin{aligned} Z_{n+1}^\tau - Z_n^\tau &= \begin{cases} Z_{n+1} - Z_n & \text{, falls } \tau \geq n+1 \\ 0 & \text{, falls } \underbrace{\tau \leq n}_{=:E \in \mathcal{J}_n} \end{cases} \\ &= 1_{\{\tau \geq n+1\}}\left(Z_{n+1} - Z_n\right) \end{aligned}$$

Hiermit gilt dann

$$\begin{aligned} \mathbb{E}\left(Z_{n+1}^\tau - Z_n^\tau \mid \mathcal{J}_n\right) &= \mathbb{E}\left(1_{\underbrace{\{\tau \geq n+1\}}_{\in \mathcal{J}_n}}\left(Z_{n+1} - Z_n\right) \mid \mathcal{J}_n\right) \\ &= 1_{\{\tau \geq n+1\}}\left[\mathbb{E}\left(Z_{n+1} - Z_n \mid \mathcal{J}_n\right) - Z_n\right] \\ &= \begin{cases} 0 & \text{, falls } \tau \leq n \\ \mathbb{E}\left(Z_{n+1} \mid \mathcal{J}_n\right) - Z_n & \text{, falls } \tau \geq n+1 \end{cases} \end{aligned}$$

Falls $\tau \geq n+1$ folgt $Z_n > X_n$ und daher gilt auf der Menge auf $\{\tau \geq n+1\}$

$$\begin{aligned} Z_n : &= \max\left(\mathbb{E}\left(Z_{n+1} \mid \mathcal{J}_n\right), X_n\right) \\ &= \mathbb{E}\left(Z_{n+1} \mid \mathcal{J}_n\right). \end{aligned}$$

Insgesamt erhalten wir also

$$\mathbb{E}\left(Z_{n+1}^{\tau} - Z_n^{\tau} \mid \mathcal{J}_n\right) = 0.$$

Folglich ist $Z_{\cdot}^{\tau}$ ein Martingal. □

Korollar 103. *τ ist eine optimale Stoppzeit, und für den Wert des Optimalen Stopproblems gilt*

$$\sup_{\sigma} E(X_\sigma) = E(X_\tau) = Z_0.$$

Beweis. Mit der Martingaleigenschaft von Z^τ schließen wir, dass

$$Z_0 = Z_0^\tau = \mathbb{E}\left(Z_N^\tau\right) = \mathbb{E}\left(Z_{N\wedge\tau}\right) = \mathbb{E}\left(Z_\tau\right) = \mathbb{E}\left(X_\tau\right).$$

Falls $\pi \leq N$ eine beliebige andere Stoppzeit ist, gilt zudem mit Optional Stopping für das Supermartngal $Z_{\cdot}$ sowie $Z_\pi \geq X$, dass

$$Z_0 \geq \mathbb{E}\left(Z_\pi\right) \geq \mathbb{E}\left(X_\pi\right).$$

Somit ist τ optimal. □

Bemerkung 104. Der vorausgegangene Satz kann grundsätzlich für eine praktische Implementierung einer optimalen Stoppregel genutzt werden. Der Einfachheit halber sei vorausgesetzt, dass $\mathcal{J}_{\cdot} = \mathcal{F}_{\cdot}^X$. Zudem sei die Verteilung von $X_{\cdot}$ bekannt. Wir können dann die Zufallsvariablen $Z_{\cdot}$ per Rückwärtsinduktion berechnen wie folgt

$$Z_N := X_N, \quad Z_{n-1} := \max\left(X_{n-1}, \mathbb{E}\left(Z_n \mid X_1, ..., X_{n-1}\right)\right).$$

Nach dem Faktoriserungslemma lässt sich Z_n insbesondere darstellen in der Form

$$Z_n := f_n(X_1, \cdots, X_n).$$

mit einer gewissen (und vielen Fällen explizit berechenbaren) Funktion $f_n : \mathbb{R}^n \to \mathbb{R}$. Mit Hilfe dieser Funktion können wir dann mit fortschreitender Zeit und bei $n = 0$ beginnend überprüfen, ob der Fall $f_n(X_1(\omega), \cdots, X_n(\omega)) = X_n(\omega)$ vorliegt oder nicht. Falls ja, so haben wir $\tau(\omega) = n$ gefunden, andernfalls gehen wir einen Zeitschritt weiter und testern erneut.

3.6. Martingalungleichungen und Konvergenzsätze.

3.6.1. *Maximalungleichungen.* Die Submartingaleigenschaft eines stochastischen Prozesses beschreibt eine wachsende Tendenz der Pfade im Hinblick auf die bedingten Erwartungen. Jedoch sind damit die Pfade nicht monoton wachsend fast sicher, wie z.b. die Summe von i.i.d. Realisierungen eines gezinkten Münzwurfs zeigt. Das folgende Lemma zeigt jedoch, dass das pfadweise Maximum eines Submartingals im Sinne einer Tschebyschev-Ungleichung durch den Endwert der Trajektorien am Beobachtungshoizont abgeschätzt werden kann. (Diese Ungleichung verallgemeinert die Kolmogorov'sche Maximalungleichung für Summen von i.i.d. Zufallsvariablen.)

Lemma 105 (Maximalungleichung für Submartingale)**.** *Sei* $(X_\cdot)$ *ein Submartingal und sei* $\overline{X_k} := \max\{X_j \mid j \leq k\}$ *der Prozess des laufenden Maximums vom* X. *Dann gilt für* $\lambda > 0$

$$\lambda \cdot P\left(\overline{X}_T > \lambda\right) \leq \mathbb{E}\left(X_T; \overline{X}_T \geq \lambda\right).$$

bzw. für $p > 1$

$$\lambda^p \cdot P\left(\overline{X}_T > \lambda\right) \leq \mathbb{E}\left((X_T^+)^p; \overline{X}_T \geq \lambda\right).$$

Kombiniert mit der trivialen Ungleichung $\mathbb{E}\left(X_T; \overline{X}_T \geq \lambda\right) \leq \mathbb{E}\left(X_T^+\right)$ erhalten wir eine handlichere Aussage, die wir gesondert festhalten wollen.

Korollar 106. *Für ein Submartingal gilt für* $\lambda > 0$

$$\lambda \cdot P\left(\overline{X}_T \geq \lambda\right) \leq \mathbb{E}\left(X_T^+\right)$$

bzw. für $p > 1$

$$\lambda^p \cdot P\left(\overline{X}_T \geq \lambda\right) \leq \mathbb{E}\left((X_T^+)^p\right).$$

Beweis von Lemma 105. Wir betrachten zunächst den Fall $p > 1$, dann folgt aus der Monotonie und Konvexität der reellen Funktion $x \to (x^+)^p = (\max(0,x))^p$, dass $(X_\cdot^+)^p$ ebenfalls ein Submartingal ist. Führen wir die Stoppzeit $\tau := \inf\{k \mid X_k \geq \lambda\} \wedge T$ ein, so gilt mit dem Optional Sampling Theorem wegen $T \geq \tau$

$$\begin{aligned}
\mathbb{E}\left((X_T^+)^p\right) &\geq \mathbb{E}\left((X_\tau^+)^p\right) \\
&= \mathbb{E}\left((X_\tau^+)^p 1_{\{\overline{X}_T < \lambda\}}\right) + \mathbb{E}\left((X_\tau^+)^p 1_{\{\overline{X}_T \geq \lambda\}}\right) \\
&\geq \mathbb{E}\left((X_\tau^+)^p 1_{\{\overline{X}_T < \lambda\}}\right) + \lambda^p P\left(\overline{X}_T \geq \lambda\right).
\end{aligned}$$

Hieraus entsteht die Aussage durch Subtraktion von $\mathbb{E}\left((X_\tau^+)^p 1_{\{\overline{X}_T < \lambda\}}\right)$. Im Fall $p = 1$ ist der Übergang zum Positivteil von $X_\cdot$ nicht nötig, ansonsten gilt dieselbe Argumentation. □

Satz 107 (Doob'sche L^p-Maximalungleichung)**.** *Es sei* $X_\cdot$ *ein nichtnegatives Supermartingal und* $p > 1$, *dann gilt*

$$\mathbb{E}\left(\overline{X}_T^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left(X_T^p\right),$$

Korollar 108. *Für ein beliebiges Submartingal gilt für* $p > 1$

$$\mathbb{E}\left(\overline{X_T^+}^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left((X_T^+)^p\right).$$

und für ein Martingal entsprechend

$$\mathbb{E}\left(M_T^*\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left(|M_T|^p\right),$$

mit dem Prozess $M_k^* := \overline{|M|}_k = \max\{|M_j|, j \leq k\}$ *des laufenden Maximums vom Absolutbetrag.*

Beweis. Die erste Aussage ergibt sich durch Anwendung des vorigen Sates auf das Submartingal $X_\cdot^+$, die zweite entsteht durch Anwendung auf $X_\cdot = |M_\cdot|$. □

Beweis von Satz 107. Wir erinnern uns zunächst noch einmal an die Methode der Berechnung von Momenten einer Zufallsvariablen $Z \geq 0$ durch Integration über Niveaumengen wie folgt

$$\begin{aligned}
\mathbb{E}\left(Z^p\right) &= \mathbb{E}\left(p\int_0^Z t^{p-1}dt\right) \\
&= p\int_\Omega\int_\mathbb{R} 1_{0\leq t\leq z} t^{p-1} dt dP \\
&= p\int_0^\infty \int_\Omega 1_{\{Z\geq t\}} dP dt \\
&= p\int_0^\infty t^{p-1} P\left(Z\geq t\right) dt.
\end{aligned}$$

Nun zum Beweis der Aussage. Sei $k > 0$, dann

$$\mathbb{E}\left(\left(\overline{X}_T \wedge k\right)^p\right) = \int_0^k p\cdot\lambda^{p-1} P\left(X_T\geq\lambda\right) d\lambda (*)$$

Eingesetzt in die vorige Ungleichung ergibt das (im Falle $p = 1$)

$$\begin{aligned}
(*) &\leq p\int_0^k \lambda^{p-2}\mathbb{E}\left(X_T 1_{\left\{\overline{X}_T>\lambda\right\}}\right) \\
&= p\mathbb{E}\left(X_T\int_0^k \lambda^{p-2} 1_{\{X_T<\lambda\}}\right) d\lambda \\
&= p\mathbb{E}\left(X_T\int_0^{\overline{X}_T\wedge k}\lambda^{p-2}d\lambda\right) \\
&= \frac{p}{p-1}\mathbb{E}\left(X_T\left(\overline{X}_T\wedge k\right)^{p-2}\right)
\end{aligned}$$

Für $k \to \infty$:

$$\begin{aligned}
\mathbb{E}\left(\overline{X}_T^p\right) &\leq \frac{p}{p-1}\mathbb{E}\left(X_T\cdot\overline{X}_T^{p-2}\right) \\
&\overset{\text{Hölder}}{\leq} \frac{p}{p-2}\left(\mathbb{E}\left(X_T^p\right)\right)^{\frac{1}{p}}\cdot\left(\mathbb{E}\left(\overline{X}_T^p\right)\right)^{\frac{p-1}{p}}
\end{aligned}$$

□

3.6.2. *Martingal-Konvergenzsatz.* Der folgende Satz bestätigt ein weiteres Mal eine gewisse Analogie von Submartingalen und monoton wachsenden Funktionen, auch wenn ein Submartingal pfadweise im Allgemeinen nicht monton wachsend ist. Er besagt, dass ein (in geeigneter Weise) nach oben beschränktes Submartingal für $t \to \infty$ pfadweise fast sicher gegen einen Limes konvergiert.

Satz 109 (Doobscher Konvergenzsatz)**.**

Falls $(X_k)_{k\in\mathbb{N}}$ ein Submartingal bzgl. $\mathcal{J}$ ist, s.d. $\sup_k \mathbb{E}\left(X_k^+\right) < \infty$, dann:

$$\exists X_\infty(\omega) = \lim_k X_k(\omega) \text{ für } P-\text{fast jedes } \omega\in\Omega.$$

Bemerkung 110. X_∞ ist i. A. eine nichtdeterministische Zufallsvariable.

Korollar 111. *Die Pfade eines nichtnegativen Martingals konvergieren fast sicher.*

Der Beweis von Satz 3.6.2 beruht auf dem sogenannten *Überkreuzungslemma* unten. Gegeben zwei reelle Zahlen $a < b$ definieren wir induktiv eine Folge von Stoppzeiten $(S_k)_{k\in\mathbb{N}}$ durch $S_1(\omega) := \inf\{k \mid X_k(\omega) > b\} \wedge T$ und $S_2(\omega) := \inf\{k \mid X_k(\omega) < a\} \wedge T,\ S_{2n+1}(\omega) = \inf\{k \geq S_{2n} \mid X_k(\omega) > a\} \wedge T, T \in \mathbb{N}$ und $S_{2n}(\omega) = \inf\{k \geq S_{2n-1} \mid X_k < b\} \wedge T$ als Folge der alternierenden Austrittzeiten des Prozesses aus dem Intervall $[a, b]$.

Die Zahl $D(\omega) = D([a, b], T)(\omega) = \sup\{n \mid S_{2n} < T\}$, gibt dann die der Anzahl der absteigenden Überquerungen des Wertebereiches $[a, b]$ im Zeitintervall $\{0, ..., T\}$ durch die Trajekorie $X_n(\omega)$ an.

Lemma 112 (Überkreuzungslemma)**.** *Falls $X.$ ein Submartingal ist, so gilt für $a < b$, $T \in \mathbb{N}$*

$$E(D([a, b], T)) \leq \frac{1}{b-a} E((X-b)^+)$$

Beweis. Zu $k \in \mathbb{N}$ definieren wir $A_k := \{S_k < T\} \in \mathcal{J}_{S_k}$, dann gilt $A_{k+1} \subset A_k, X_{s_{2k-1}} \geq b$ auf A_{2k-1} und $X_{s_{2k}} \leq a$ auf A_{2k}.

$$\begin{aligned}
\Longrightarrow 0 \quad &\leq \quad \int_{A_{2k-1}} \left(X_{s_{2k-1}} - b\right) dP \\
&\overset{\text{opt.Sam.}}{\leq} \quad \int_{A_{2k-1}} (X_{s_{2k}} - b)\, dP \\
&= \quad \int_{A_{2k}} (X_{s_{2k}} - b)\, dP + \int_{A_{2k-1}\setminus A_{2k}} (X_{s_{2k}} - b)\, dP \\
&\leq \quad \int_{A_{2k}} \underbrace{(a-b)}_{\leq 0}\, dP + \int_{A_{2k-1}\setminus A_{2k}} (X_{s_{2k}} - b)^+\, dP
\end{aligned}$$

Folglich

$$(b-a)P(A_{2k}) \leq \int_{A_{2k-1}\setminus A_{2k}} (X_{s_{2k}} - b)^+\, dP$$

A_{2k} entspricht $\{D \geq k\}$, d.h. $P(A_{2k}) = P(D \geq k)$

$$\begin{aligned}
\Longrightarrow (b-a)\sum_{k=0}^{\infty} P(D \geq k) \quad &\leq \quad \sum_{k=1}^{\infty} \int_{A_{2k-1}\setminus A_{2k}} (X_{s_{2k}} - b)^+\, dP \\
&\overset{\text{Subm.}}{\leq} \quad \sum_{k=0}^{\infty} \int_{A_{2k-1}\setminus A-2k} (X_T - b)^+\, dP = (*)
\end{aligned}$$

$A_{2k-1}\setminus A_{2k} \cap A_{2j-1}\setminus A_{2j} = \varnothing$ für $k \neq j$.

$$\begin{aligned}
\Longrightarrow (*) \quad &= \quad \int_{\bigcup_{k\in\mathbb{N}_0} A_{2k-1}\setminus A_{2k}} (X_T - b)^+\, dP \leq \int_\Omega (X_T - b)^+\, dP \\
&= \quad \mathbb{E}\left((X_T - b)^+\right)
\end{aligned}$$

$$\Longrightarrow (b-a)\mathbb{E}(D) \leq \mathbb{E}\left((X_T - b)^+\right)$$

□

Beweis des Doob'schen Konvergenzsatzes. Falls X_k ein Submartingal ist, so auch X_k^+ Submartingal bzw. ebenso $(X_k - b)^+$.

$$\begin{aligned} \Rightarrow \sup_{k\in\mathbb{N}} \mathbb{E}\left((X_k - b)^+\right) &= \lim_k \mathbb{E}\left((X_k - b)^+\right) \\ &\leq b + \lim_k \mathbb{E}\left(X_k^+\right) < \infty \end{aligned}$$

d.h.

$$\sup_k \mathbb{E}\left((X_k - b)^+\right) < \infty$$

Sei $\overline{X(\omega)} = \limsup_k X_k(\omega)$. Beh.: $\overline{X(\omega)} = \lim_k X_k(\omega)$ für $P-$fast alle $\omega \in \Omega$. Denn andernfalls existieren (a,b), s.d. $[a,b]$ unendlich häufig absteigend überquert wird. Somit gilt die Inklusion

$$\left\{\omega \mid \overline{X(\omega)} \neq \lim_k X_k(\omega)\right\} \subset \bigcup_{a,b\in\mathbb{Q},a<b} \{D([a,b],\infty) = \infty\}$$

wobei

$$D\left([a,b],\infty\right) = \lim_T D\left([a,b],T\right).$$

Mit monotoner Konvergenz und dem Überkreuzungslemma folgt

$$\begin{aligned} \mathbb{E}\left(D([a,b],\infty)\right) &\leq \lim_T \mathbb{E}\left(D\left([a,b],\infty\right)\right) \\ &\leq \lim_T \frac{1}{b-a}\mathbb{E}\left((X_T - b)^+\right) \\ &\leq \frac{1}{b-a}\left(-b + \sup_k \mathbb{E}\left(X_k^+\right)\right) < \infty. \end{aligned}$$

Somit ist $\{D\left([a,b],\infty\right) = \infty\}$ eine P-Nullmenge, also auch $\bigcup_{a<b,a,b\in\mathbb{Q}}\{D\left([a,b],\infty\right) = \infty\}$ eine P-Nullmenge als abzählbare Vereinigung von Nullmengen. $\square$

3.7. **Anwendung: Galton-Watson-Prozess.** Der Galton-Watson Prozess ist das klassische Modell der Wahrscheinlichkeitstheorie zur Beschreibung eines Familienstammbaums, wobei der Einfachheit davon ausgegangen wird, dass Individuen sich unabhängig voneinander und ohne Zutun Dritter (etwa durch Zellteilung o.ä.) vermehren. Zudem wird vorausgesetzt, dass ein Individuum mit der Geburt seiner Nachkommen abstirbt.

Z_n sei die Zufallsvariable, die die Anzahl der Individuen der n-ten Generation, und damit die Größe der Gesamtpopulation zum Zeitpunkt n beschreibt. Da wir die Nachkommenschaft eines einzelnen Individuums bei Start in $t = 0$ beschreiben wollen, setzen wir $Z_0 = 1$. Zur Modellierung der weiteren Nachkommenschaft sei $\left(\left\{\xi_j^n\right\}_{j\in\mathbb{N}}\right)_{n\in\mathbb{N}} \in \mathbb{N}_0$ eine Doppelfolge von i.i.d. $\mathbb{N}_0$-wertigen Zufallsvariablen. Wir interpretieren ξ_j^{n+1} als Anzahl der Nachkommen des Individuums j, welches aus der n-ten Generation stammt. Wir definieren hiermit dann induktiv

$$\begin{aligned} Z_1 &= \xi_1^1 \\ Z_2 &= \sum_{j=1}^{\xi_1^1} \xi_1^2 \\ &\cdots \\ Z_{n+1} &= \sum_{j=1}^{Z_n} \xi_j^{n+1}. \end{aligned}$$

Der Erwartungswert

$$\mu = \mathbb{E}\left(\xi_1^1\right)$$

ist die mittlere Kinderzahl pro Individuum. Je nachdem, ob $\mu < 1$, $\mu = 1$ oder $\mu > 1$ spricht man vom *subkritischen, kritischen bzw. superkritischen Fall.*

Wir definieren

$$\mathcal{J}_n = \sigma\left(\xi_k^j, k \in \mathbb{N}, j \in \{1, ..., n\}\right)$$

als die σ-Algebra, die von $\left\{\xi_j^k\right\}$ bis zur n-ten Generation erzeugt wird. Offenbar ist Z_n messbar bzgl. $\mathcal{J}_n$.

$$\begin{aligned} \mathbb{E}\left(Z_{n+1} \mid \mathcal{J}_n\right) &= \mathbb{E}\left(\sum_{j=1}^{Z_n} \xi_j^{n+1} \mid \mathcal{J}_n\right) \\ &= \sum_{j=1}^{Z_n} \underbrace{\mathbb{E}\left(\xi_j^{n+1} \mid \mathcal{J}_n\right)}_{=\mu} \\ &= Z_n \cdot \mu \end{aligned}$$

Aus der Behauptung folgt mit vollständiger Induktion, dass

$$X_n := \frac{Z_n}{\mu^n}$$

ein nichtnegatives $(\mathcal{J}., P)$-Martingal darstellt mit

$$\mathbb{E}\left(X_n\right) = \mathbb{E}\left(X_n^+\right) = 1 < \infty.$$

Aus dem Doob'schen Konvergenzseatz folgern wir die fast sichere Existenz des Grenzwertes

$$\exists X_\infty := \lim_{n\to\infty} X_n \text{ f.s.}$$

d.h. falls $X_\infty(\omega) = \lambda(\omega)$, dann

$$X_k(\omega) \approx \mu^k \lambda(\omega)$$

für große k. Fast sicher folgt die Entwicklung der Familiengröße also asymptotisch einem exponentiellen Wachstumsgesetz, wobei die Rate λ zufällig ist.

Wir bertachten nun die Frage nach dem Aussterben der Gesamtpopulation. Sei hierzu $\xi := \inf\{n \in \mathbb{N}_0 \mid Z_n = 0\}$ der Zeitpunkt des Aussterbens der Familie

Satz 113. *Falls $\mu < 1$, dann ist $\xi < \infty$ fast sicher.*

Beweis. $X_n := \frac{Z_n}{\mu_n}$ ist Martingal, somit $\mathbb{E}(Z_n) = \mu^n$. Ferner gilt offensichtlich, dass $Z_n \geq 1$ auf $\{Z_n > 0\}$. Daher

$$\begin{aligned} P(Z_n > 0) &\leq \mathbb{E}(Z_n; Z_n > 0) \\ &= \mathbb{E}(Z_n) \\ &= \mu^n \end{aligned}$$

Wegen $\mu < 1$ schließen wir, dass $\sum_{n \in \mathbb{N}_0} P(Z_n > 0) < \infty$. Also folgt nach Borel-Cantelli

$$P(Z_n > 0 \text{ für unendlich viele } n) = 0.$$

□

Satz 114. *Falls $\mu = 1$ und $\mathbb{V}(\xi_1^1) > 0$ so ist $\xi < \infty$ fast sicher.*

Beweis. Für den folgenden Beweis benutzen wir, dass $Z.$ eine Markov-Kette auf $\mathbb{N}_0$ ist, wie man leicht an der rekursiven Konstruktion von $Z.$ erkennt. Als weitere Vorbereitung macht man sich klar, dass für $k \in \mathbb{N}$ (d.h $k \neq 0$)

$$\gamma := P(Z_{N+M} = k \mid Z_{N+M-1} = k) < 1,$$

denn der Fall $\gamma = 1$ wäre wegen $E(\xi_1^1) = 1$ nur für determische ξ_j^i möglich, d.h. wenn $\mathbb{V}(\xi_1^1) = 0$.

Hier nun das Argument. Wegen der Ganzzahligkeit von $Z.$ impliziert die Konvergenz, dass $n \to Z_n(\omega)$ schließlich fast sicher konstant ist. Für die Identifikation des Limes seien $N, M \in \mathbb{N}$, $k \in \mathbb{N}$, dann

$$\begin{aligned} P(Z_n = &k; \forall n \geq N) \leq P(Z_n = k; \forall n \in \{N, ..., N+M\}) \\ &= P(Z_{N+M} = k \mid Z_N = k, ..., Z_{N+M-1} = k)\, P(Z_N = k, ..., Z_{N+M-1} = k) \\ &\overset{Z.\text{ Markov}}{=} \underbrace{P(Z_{N+M} = k \mid Z_{N+M-1} = k)}_{\gamma<1} P(Z_n = k \mid n \in \{N, ..., N+M-1\}) \end{aligned}$$

und durch Iteration dieses Argumentes schließlich

$$\leq \underbrace{\gamma^M}_{\to 0} P(Z_n = k) \longrightarrow \qquad \text{falls } M \to \infty.$$

Für alle $k \in \mathbb{N}$, $N \in \mathbb{N}$ ist also

$$P(Z_n = k \forall n \geq N) = 0$$

Damit kann kommt nur noch $0 \in \mathbb{N}_0$ als konstanter Wert in Frage, auf den Sich $Z.$ pfadweise schließlich stabilisiert. Inbsesondere folgt $\xi < \infty$ f.s. □

Für den superkritischen Fall $\mu > 1$ treffen wir einige Vorbereitungen.

Definition 115. Es bezeichne $p_k := P(\xi_1^1 = k)$, $k \in \mathbb{N}_0$. Dann heißt $\varphi : [0,1] \to [0,1]$ mit

$$\varphi(s) = \mathbb{E}\left(s^{\xi_1^1}\right) = \sum_{k>0} s^k p_k$$

die *(Momenten-)erzeugende Funktion* von ξ_1^1.

Die erzeugende Funktion hat (im Fall $\mu > 1$) die folgenden Eigenschaften.

$$\begin{aligned}
\varphi(0) &= p_0 < 1 \\
\varphi'(s) &= \underbrace{\sum_{k\geq 1} k \cdot s^{k-1} p_k}_{>0} \Longrightarrow \varphi'(1) = \mathbb{E}\left(\xi_1^1\right) = \mu \\
\varphi''(s) &= \sum_{k\geq 2} k(k-1) s^{k-2} p_k > 0 \Longrightarrow \varphi \text{ konvex}
\end{aligned}$$

Insbesondere hat $s \mapsto \varphi(s)$ einen eindeutigen Fixpunkt ρ in $[0,1[$, d.h. Lösung der Gleichung $\varphi(\rho) = \rho$ (man mache sich eine Skizze) und für die rekursiv definierte Folge

$$a_0 = p_0, \quad a_{n+1} := \varphi(a_n)$$

gilt

$$\lim_{n\to\infty} a_n = \rho.$$

Satz 116. *Für den Galton-Watson-Prozess gilt im Fall $\mu > 1$, dass*

$$P(\xi < \infty) = \rho.$$

Beweis. Für $n \in \mathbb{N}_0$ sei $\theta_n := P(\xi \leq n) = P(Z_n = 0)$. Dann ist $\theta_0 = p_0$. Für $n \geq 1$ argumentieren wir wie folgt. Falls $Z_1 = k$, so ist die Gesamtpopulation zum Zeitpunkt n genau dann ausgestorben, wenn sämtliche von den jeweiligen Mitgliedern der ersten Generation abstammenden Teilfamilien im restlich verbleibenden Zeitraum $n-1$ ausgestorben sind. Da die weitere Entwicklung der jeweiligen Nachkommenschaften von derselben Generation stochastisch unabhängig sind, gilt also

$$P(Z_n = 0 | Z_1 = k) = \theta_{n-1}^k.$$

Insgesamt erhalten wir damit

$$P(Z_n = 0) = \sum_{k\geq 0} P(Z_n = 0|Z_1 = k) P(Z_1 = k) = \sum_{k\geq 0} \theta_{n-1}^k p_k = \varphi(\theta_{n-1}).$$

Die Folge (θ_n) folgt daher dem oben beschriebenen Rekursionsgesetz, somit ist $P(\xi < \infty) = \lim_n \theta_n = \rho$. □

3.8. **Abschließbarkeit von Martingalen.** Es stellt sich die Frage, ob man durch Hinzunahme des pfadweisen Grenzwertes X_∞ ein (Sub-)Martingal zu einem (Sub-)Martingal $(X_k)_{l\in\mathbb{N}_0\cup\{\infty\}}$ fortsetzen kann.

Definition 117. Ein Martingal $(X_k)_{k\in I}$ heißt abschließbar, falls eine $\mathcal{J}_\infty := \sigma\left(\{\mathcal{J}_k \mid k \in I\right)$-messbare Zufallsvariable X_∞ existiert, so dass für alle $t \in \mathbb{N}_0$ $X_t = \mathbb{E}_p\left(X_\infty \mid \mathcal{J}_t\right)$.

Die Frage nach der Abschließbarkeit tritt z.B. auf, wenn man das Optional Sampling für Martigale bei unbeschränkte Stoppzeiten anwenden kann.

Satz 118. *Es sei $(X_k)_{k\in\mathbb{N}_0}$ ein abschließbares Martingal und $S \leq T : \Omega \to \mathbb{N}_0$ zwei Stoppzeiten, dann gilt $E(X_T|\mathcal{J}_S) = X_S$ fast sicher.*

Beweis. Analog zu den ersten beiden Beweisschritten vom Optional Sampling Theorem zeigt man zunächst, dass $X_T = E(X_\infty|\mathcal{J}_T)$ und benutzt dann die Turmeigenschaft der bedingten Erwartung. □

Bemerkung 119. Satz 118 bleibt richtig für $\mathbb{N}_0 \cup \{\infty\}$-wertige Stoppzeiten.

Das entscheidende Kriterium für die Abschließbarkeit von Martingalen ist das folgende.

Definition 120. Eine Familie von reellen Zufallsvariablen $(X_k)_{i\in I}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ heißt *gleichgradig integrierbar*, falls

$$\sup_{i\in I} \int_{\{|X_i|>c\}} |X_i| dP \overset{c\to\infty}{\longrightarrow} 0.$$

Die gleichgradige Integrierbarkeit einer Familie von Zufallsvariablen ist im wesentlichen Äquivalent zu deren Präkompaktheit im Sinne der L^1-Konvergenz.

Satz 121 (Vitalischer Konvergenzsatz)**.**

$$f_k \to f \;\; in \;\; L^1(\Omega, f, P) \iff \begin{cases} \{f_k\} & \textit{gleichgr. int'bar und} \\ f_k \to f & \textit{stochastisch} \end{cases}$$

bzw. in der L^p-Version gilt

$$f_k \to f \;\; in \;\; L^p \iff \begin{cases} \{|f_k|^p\}_{k\in\mathbb{N}} & \textit{gleichgr. int'bar und} \\ f_k \to f & \textit{stochastisch.} \end{cases}$$

Der Lebesgue'sche Satz von der dominierten Konvergenz ergibt sich als Korollar zum Vitalischen Konvergenzsatz aus dem folgenden Lemma, dessen Beweis wir als Übungsaufgabe stellen (vergleiche hierzu den Beweis von Lemma 124).

Lemma 122. *Es sei $\{X_k\}$ eine Familie von reellen Zufallsvariablen und $Z \geq 0$ eine Zufallsvariable mit $E(Z) < \infty$, so dass $|X_k| \leq Z$ fast sicher, dann ist die Familie $\{X_k\}$ gleichgradig integrierbar.*

Vor dem Beweis des Vitali'schen Konvergenzsatzes wollen wir noch eine nützliche äquivalente Charakterisierung der gleichgradigen Integrierbarkeit bereit stellen.

Lemma 123. *Die Familie $(X_i)_{i\in I}$ ist gleichgr. int'bar ganau dann, wenn*

$$\sup_{i\in I} E(|X_i|) < \infty$$

und

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall A \in \mathcal{J} : P(A) < \delta \Longrightarrow \int_A |X_i| dP < \varepsilon, \forall i \in I.$$

Den Beweis hiervon stellen wir als Übungsaufgabe. Es handelt sich dabei im Wesentlichen um eine Modifikation vom Beweis der nächsten Aussage.

Lemma 124. *Falls $Y \in L^1(\Omega, P)$, dann ist $\{Y\}$ (als einelementige Menge von Zufallsvariablen) gleichgradig integrierbar.*

Beweis. Wir zeigen, dass die L^1-Integrierbarkeit die in Lemma 123 genannten Kriterien nach sich zieht. Das erste Kriterium ist trivialer Weise erfüllt . Für das zweite benutzen wir, dass für eine beliebige messbare Menge A und $c > 0$ gilt

$$\begin{aligned} \int_A Y dP &= \int_{A\cap\{|Y|<c\}} Y dP + \int_{A\cap\{|Y|>c\}} Y dP \\ &\leq cP(A) + \int_{\{|Y|>c\}} Y dP \end{aligned}$$

Führen wir das Maß $\nu(d\omega) = |Y|(\omega)P(d\omega)$ auf $(\Omega, \mathcal{J})$ ein, d.h. $\nu(B) = \int_B |Y|(\omega)P(d\omega)$, so ist dies wegen $\nu(\Omega) = \int_\Omega |Y|dP < \infty$ endlich. Inbesondere gilt die 'Stetigkeit von Oben' von Maßen für das endliche Maß ν.

Für $c > 0$ sei $A_c := \{|Y| > c\}$. Dann ist $A_\infty := \bigcap_{c>0} A_c = \{|Y| = \infty\}$ eine P-Nullmenge wegen $Y \in L^1(P)$. Daher ist A_∞ auch eine ν-Nullmenge, da

$$\nu(A_\infty) = \int_{A_\infty} |Y|dP = 0.$$

Aus der Stetigkeit von Oben für ν folgt aus $A_c \searrow A_\infty$, dass $\lim_{c\to\infty} \nu(A_c) = 0$, d.h. zu $\epsilon > 0$ existiert ein c_ϵ, so dass $\nu(A_c) \leq \varepsilon/2$ falls $c > c_\varepsilon$. Ausgeschrieben bedeutet das

$$\int_{\{|Y|>c\}} |Y|dP < \frac{\varepsilon}{2}$$

Sei nun $\rho := \frac{\varepsilon}{2c}$, dann gilt für alle $A \in \mathcal{F}, P(A) < \rho$

$$\int_A |Y|dP \leq cP(A) + \int_{\{|Y|>c\}} |Y|dP < c\frac{\varepsilon}{2c} + \frac{\varepsilon}{2} = \varepsilon.$$

□

Beweis des Vitali'schen Konvergenzsates. "$\Rightarrow$" Falls $X_n \to X$ in L^1, so auch stochastisch wegen Tschebbyschev. Weiter ist

$$\begin{aligned}\int_{\{|X_n|>c\}} |X_n|dP &\leq \int_{\{|X_n|>c\}} |X - X_n| + \int_{\{|X_n|>c\}} |X|dP \\ &\leq ||X - X_n||_{L^1} + \int_{\{|X_n|>c\}} |X|dP\end{aligned}$$

mit $A = A_n(c)$ gilt gilt

$$P(A_n) \leq \frac{1}{c}\mathbb{E}(|X_n|) \overset{n\to\infty}{\longrightarrow} \frac{1}{c}\mathbb{E}(X),$$

so dass

$$P(A_n) < \delta \text{ für alle } n \text{ falls , falls } c > c_\delta$$

Somit

$$\int_{A_n} |X_n|dP \leq ||X - X_n||_{L^1} + \int_{A_n} |X|dP$$

$$\Longrightarrow \lim_{c\to\infty} \sup_n \int_\sim |X_n|dP \leq \limsup_n ||X - X_n||_{L^1} + \lim_{c\to\infty} \sup_{A\in\mathcal{J}, P(A)<\frac{1}{c}} \int_A |X|dP = 0,$$

webei wir im letzten Schritt Lemma 124 ausgenutzt haben.

"$\Leftarrow$" Wir betrachten zunächst den Spezialfall, dass ein $c > 0$ existiert mit $(X_k) \leq c \forall k$. OBdA sei $X = 0$ und $X_k \to 0$ stochastisch. Wir müssen zeigen, dass

$$\|X_k\| = \int X_k^+ dP + \int X_k^- dP \to 0 \text{ für } k \to \infty,$$

wobei $X^+ := \max(0, X)$ und $X^- := (-X)^+$ den Positiv- und Negativteil einer Zufallsvariablen X bezeichnen. Mit X_k konvergieren auch X^+ und X^-

stochastisch gegen Null, so dass es ausreicht, den Fall $X_k \geq 0$ zu betrachten mit $X_k \leq c \forall k$ für ein gewisses $c > 0$. Sei nun $\epsilon > 0$, dann gilt

$$\begin{aligned}\int X_k dP &= \int_{\{X_k \geq \epsilon\}} X_k dP + \int_{\{X<\epsilon\}} X_k dP \\ &\leq cP(X_k \geq \epsilon) + \epsilon,\end{aligned}$$

d.h. aufgrund der stochastischen Konvergenz von $X_k \to 0$

$$\limsup_{k\to\infty} \int X_k dP \leq \epsilon,$$

mit $\epsilon > 0$ beliebig, also die Behauptung.

$X_n \to X$ in L^1 mittels dominierter Konvergenz.
Im allgemeinen Fall gilt für $c > 0$, dass auch die Folge der $\overline{X}_k = X_k \cdot 1_{\{|X_k| \leq c\}}$ stochastisch gegen $\overline{X} := X \cdot 1_{\{|X| \leq c\}}$ konvergiert, so dass

$$\overline{X}_k \longrightarrow \overline{X} \text{ in } L^1.$$

Die Aussage ergibt sich dann aus der Zerlegung

$$||X_k - X||_{L^1} \leq ||X_k - \overline{X}_k||_{L^1} + ||\overline{X}_k - \overline{X}||_{L^1} + ||\overline{X} - X||_{L^1} \to 0$$

und

$$\sup_{k\in\mathbb{N}} \|\overline{X}_k - X_k\| = \sup_{k\in\mathbb{N}} \int_{\{|X_k| \geq c\}} |X_k| dP \to 0 \text{ für } c \to 0,$$

da die Folge (X_k) nach Voraussetzung gleichgradig integrierbar ist. □

Nach diesen Vorbereitungen können wir das Hauptergebnis dieses Abshnitts formulieren.

Satz 125 (Abschließbarkeit von Martingalen). *Sei $\{X_k\}$ ein Martingal. Dann sind äquivalent*

(1) $\{X_k\}$ *ist gleichgradig integrierbar.*
(2) $\exists X_\infty$, *s.d.* $X_k = \mathbb{E}(X_\infty \mid \mathcal{J}_k), X_\infty \in L^1$.
(3) $\exists \lim_{k\to\infty} X_k =: Z$ *im Sinne der L^1-Konvergenz.*

Bemerkung 126. Insbesondere ist jedes beschränkte Martingal abschließbar.

Beweis. $1 \Rightarrow 3$. Sei $\{X_k\}$ gleichgradig integrierbar

$$\Longrightarrow \sup_{k\in\mathbb{N}_0} \mathbb{E}(|X_k|) < \infty \Longrightarrow \sup_k \mathbb{E}\left(X_k^+\right) < \infty.$$

Nach dem Doob'schen Konvergenzsatz existiert fast sicher $Z = \lim_{k\to\infty} X_k$, also auch $X_k \to Z$ stochastisch. Da $(X_k)_{k\in\mathbb{N}_0}$ glgr. intbar, folgt 3. mit Vitali.
$3 \Rightarrow 1$ ist die Rückrichtung von Vitali.
$2 \Rightarrow 3$. Falls 2. gilt, dann ist $(X_k)_{k\in\mathbb{N}_0\cup\{\infty\}}$ ein Martingal und $(|X_k|)_.$ ein Submartingal. Somit

$$\int_{\underbrace{|X_k| > c}_{\in\mathcal{J}_k}} |X_k| dP \overset{\text{Subm.}}{\leq} \int_{\{|X_k|>c\}=:A_k} |X_\infty| dP,$$

wobei

$$P(A_k) \leq \frac{1}{c}\mathbb{E}(|X_k|) \leq \frac{1}{c}\mathbb{E}(|X_\infty|) \leq \delta$$

unabhängig von k, falls $c > 0$ hinreichend groß gewählt. Gemäß Lemma 124 gilt

$$\forall \varepsilon > 0 \exists \delta > 0 : \int_A |X_\infty| dP < \varepsilon \text{ falls } \delta > P(A)$$

so dass wir insgesamt schließen, dass

$$\forall \varepsilon > 0 \exists c > 0 : \sup_k \int_{\{|X_k| > c\}} |X_k| dP \le \int_{A_k} |X_\infty| dP < \varepsilon \text{ falls } c > \frac{\mathbb{E}\left(|X_\infty|\right)}{\delta}$$

$3 \Rightarrow 2.$

$$\forall A \in \mathcal{J}_k, l \ge k : \int_A X_l dP = \int_A X_k dP$$

Für $l \to \infty$ und 3.:

$$\int_A X_\infty dP = \int_A X_k dP \Longrightarrow X_k = \mathbb{E}\left(X_\infty \mid \mathcal{J}_k\right)$$

□

Zum Schluss dieses Abschnitts geben wir noch die folgende L^p-Variante des vorigen Satzes an.

Satz 127. *Sei $(X_k)_{k \in \mathbb{N}_0}$ ein Martingal. Dann sind äquivalent:*

(1) $\sup_{k \in \mathbb{N}_0} \mathbb{E}\left(|X_k|^p\right) < \infty$.
(2) $X_\infty = \lim X_k$ *existiert als Limes in* L^p.

Beweis. $1 \Rightarrow 2$. Nach der Doob'schen Maximalungleichung folgt für

$$Y := \sup_{k \in \mathbb{N}} |X_k| \Rightarrow \mathbb{E}\left(|Y|^p\right) \le c(p) \underbrace{\sup_k \mathbb{E}\left(|X_k|^p\right)}_{< \infty}$$

Dann ist $\{|X_k|^p, k \in \mathbb{N}\}$ gleichgardig integrierbar und $X_k \to X_\infty$ f.s. Daraus folgt mit der L^p-Variante von Vitali

$$X_k \longrightarrow X_\infty \text{ in } L^p.$$

□

3.9. Anwendung: Der Satz von Radon-Nikodym.

Definition 128. Es sei $(\Omega, \mathcal{J})$ ein messbarer Raum und Q, P zwei Maße auf $(\Omega, \mathcal{J})$. Dann heißt Q *absolut stetig gegenüber* P (Notation $Q \ll P$), falls für alle $A \in \mathcal{J}$ gilt

$$P(A) = 0 \Longrightarrow Q(A) = 0.$$

Satz 129 (Radon-Nikodym)**.** *Für zwei Wahrscheinhlichkeitsmaße Q, P auf einem messbaren Raum $(\Omega, \mathcal{J})$ gilt $Q \ll P$ genau dann, wenn ein $0 \le X \in L^1(\Omega, \mathcal{J}, Q)$ existiert, so dass*

$$Q(A) = \int_A X(\omega) P(d\omega) \quad \forall\, A \in \mathcal{J}.$$

Bemerkung 130. Die Zufallsvarible X heißt die *Radon-Nikodym-Dichte von Q bzgl. P*. Schreibweise $X = \frac{dQ}{DP}$.

Der Satz von Radon-Nikodym verallgemeinert den Hauptsatz der Differential- und Integralrechnung.

Beispiel 131 (Hauptsatz der Differentialrechnung)**.** Es sei $(\Omega, \mathcal{J}) = ([0,1], \mathcal{B}[0,1])$ und $F : [0,1] \to [0,1]$ mit $F(0) = 0$ und $F(1) = 1$ die Verteilungsfunktion eines Wahrscheinhlichkeitsmaßes Q auf $[0,1]$, d.h. $Q((a,b]) = F(b) - F(a)$. Es gilt dann $Q \ll \lambda$ (absolute Stetigkeit gegenüber dem Lebesgue-Maß, wenn ein $\varphi \in L^1([0,1], \lambda)$ existiert, so dass

$$F(b) - F(a) = \int_a^b \varphi(s)\lambda(ds)$$

$$\Longrightarrow \frac{dQ}{d\lambda}(s) = \varphi(s) = F'(s).$$

Vor dem Beweis des Satzes von Radon-Nikodym zeigen wir zunächst eine Hilfsbehauptung, wonach die absolute Stetigkeit von Q bzgl. P eine Stetigkeit vom ϵ-δ-Typ nach sich zieht.

Lemma 132. *Falls $Q \leq P$, so gilt*

$$\forall \varepsilon > 0 \exists \delta > 0 : P(A) < \delta \Longrightarrow Q(A) < \varepsilon$$

Beweis. Angenommen es ex. $\varepsilon > 0 : \forall \delta > 0 : \exists A : P(A) < \delta, Q(A) > \delta$. Wähle $\delta = S_n := \left(\frac{1}{2}\right)^n$. Dann $A_n \in \mathcal{J} : P(A) < \left(\frac{1}{2}\right)^n, Q(A) > \varepsilon$. Sei $\overline{A} = \limsup_{n\to\infty} A_n, \overline{A}_k = \bigcup_{j \geq k} A_j$. Dann $\overline{A}_k \searrow \overline{A}$ für $k \to \infty$, $Q\left(\overline{A}_k\right) > \varepsilon$. Mit St. des Maßes folgt $Q(\overline{A}) > \varepsilon$

$$\sum_k P(A_k) \leq \sum_k \left(\frac{1}{2}\right)^k \overset{Bor.-Can.}{\Longrightarrow} P\left(\overline{A}\right) = 0$$

und $Q(\overline{A}) > 0$, was $Q < P$ widerspricht. □

Beweis des Satzes von Radon-Nikodym. Der Beweis beruht auf einer Anwendung der Abschließbarkeit eines geeigneten Martingals. Wir definieren hierzu die Menge aller endlichen disjunkten Zerlegungen von Ω

$$I := \{t = \{B_1, ..., B_N\} \subset 2^\Omega \,|\, \Omega = \dot\cup_{k=1}^N B_k, N \in \mathbb{N}\}$$

als Indexmenge und führen hierauf eine Ordnungsrelation ein durch

$$s \leq t \Leftrightarrow \text{ t ist Verfeinerung von } s.$$

Hiermit ist I eine (nicht total) geordnete Menge. Zur Definition einer durch I indizierten Filtration sei

$$\mathcal{F}_t := \sigma(t), \quad t \in I$$

die von t erzeugte σ-Algebra. Wir definieren

$$X_t : \Omega \to \mathbb{R}_{\geq 0}, \quad X_t(\omega) = \sum_{i=1}^{t_N} \frac{Q(B_i)}{P(B_i)} 1_{B_i}$$

mit der Konvention, dass $\frac{Q}{P}(B_i) := 0$, falls $P(B_i) = 0$.

Wir behaupten, dass $(X_t)_{t\in I}$ ein $(\mathcal{F}_t)_{t\in I}$-Martingal ist. In der Tat, betrachte $s \leq t, s = \{C_j \mid j = 1, ..., N_s\}$, dann

$$\begin{aligned}
\int_{C_j} X_t dP &= \sum_{\substack{B_i \in t \\ B_i \subset C_j}} \int_{B_i} X_t dP \\
&= \sum_{B_i \subset C_j} \int_{B_i} \frac{Q(B_i)}{P(B_i)} 1_{B_i} dP \\
&= \sum_{B_i \subset C_j} Q(B_i) \\
&= Q(C_j) \\
&= \int_{C_j} \frac{Q(C_j)}{P(C_j)} 1_{C_j} dP \\
&= \int_{C_j} X_s dP
\end{aligned}$$

Somit ist ist $I \to t \to X_t$ ein nichtnegatives Martingal mit

$$t \to \mathbb{E}\left(X_t^+\right) = \mathbb{E}(X_t) = Q(\Omega) = 1$$

Nach dem Doob'schen Konvergenzsatz existiert dann der Grenzwert $X = \lim_{t\to\infty} X_t$ P-fast sicher, wobei der Limes hier den Grenzwert entlang einer beliebigen Teilfolge $t_n \subset t_{n+1}$ $n \in \mathbb{N}$, mit $\limsup_{n\to\infty} t_n = \mathcal{F}$ bezeichnet[2].

---

[2](Konvergenz einer durch I parametrisierten Familie von Elementen eines topologischen Raumes.) Die Ordnungsrelation $\leq$ auf I ist nicht 'total' in dem Sinne, dass zwei Elemente $s, t \in I$ nicht notwendiger Weise vergleichbar sein müssen, d.h. der Fall eintreten kann, dass weder $s \leq t$ noch $t \leq s$. Für alle s, t existiert jedoch eine Majorante $s \vee t \in I$

$$s \vee t := \{B_i \cap C_j \mid B_i \in t, C_j \in s\},$$

(gemeinsame Verfeinerung), für welche $s \leq s \vee t$ und $t \leq s \vee t$. gilt.
Für eine Familie durch I parametrisierte Familie $(\alpha_t)_{t\in I}$ von Elementen eines topologischen Raumes $\alpha_t \in E$ sagen wir, dass $\lim_{t\to\infty} \alpha_t = c$ genau dann, wenn für jede folge t_k mit $t_k \nearrow \infty$ in I gilt, dass

$$\lim_{k\to\infty} \alpha_{t_k} = c.$$

Hierbei meint $t_k \nearrow \infty$ für eine Folge $(t_k)_{k\in\mathbb{N}}$ von Elementen $t_k \in I$, dass

$$t_k \leq t_{k+1}\, \forall k \in \mathbb{N} \text{ und } \forall s \in I : s \leq t_k \text{ für schließlich alle } k \in \mathbb{N}.$$

Hinreichend für die Existenz von $\lim_{t\to\infty} \alpha_t$ sind dann die beiden Bedingungen, dass

(1) $\lim_{k\to\infty} \alpha_{t_k}$ existiert für jede Folge $t_k \nearrow \infty$ und
(2) Für $t'_k \nearrow \infty$ und $t_k` \leq t_k \forall k$, dann gilt

$$\lim_{k\to\infty} \alpha_{t_k} = \lim_{k\to\infty} \alpha_{t`_k}.$$

Denn für zwei beliebige Folgen $t_k \nearrow \infty$ und $s_k \nearrow \infty$ folgt, dass $\lim_{k\to\infty} \alpha_{t_k} = \lim_{k\to\infty} \alpha_{s_k}$ durch Anwendung von (2) auf den Fall $s_k \leq s_k \vee t_k$ bzw. den Fall $t_k \leq s_k \vee t_k$.

Wir behaupten, dass die Familie $(X_t)_{t\in I}$ gleichgradig P-integrierbar ist. In der Tat

$$\begin{aligned}\int_{|X_t|\geq\alpha}|X_t|dP &= \int_{\{|X_t|\geq\alpha\}}X_tdP\\ &= \sum_{\substack{B_j\in t\\ X_t\geq\alpha}}\frac{Q(B_j)}{P(B_j)}P(B_j)\\ &= Q(\{X_t\geq\alpha\}),\end{aligned}$$

und für $\alpha>0$

$$\begin{aligned}P(\{X_t\geq\alpha\}) &\leq \frac{1}{\alpha}\mathbb{E}_p(X_t)\\ &= \frac{1}{\alpha}Q(\Omega)\\ &= \frac{1}{\alpha}.\end{aligned}$$

Mit Lemma 132 folgt, dass $\sup_{t\in I}\int_{|X_t|\geq\alpha}|X_t|dP\leq\epsilon$ falls $\alpha>1/\delta$. Somit ist $(X_t,\mathcal{J}_t)_{t\in I}$ ein abschließbares Martingal und $X_t=\mathbb{E}_p(X\mid\mathcal{J}_t)$.

Sei nun $A\in\mathcal{J}$, dann definiere $t_A:=\{A,A^c\}\in I$

$$\Longrightarrow X_{t_A}=\mathbb{E}_p(X\mid\mathcal{J}_t)$$
$$\Longrightarrow Q(A)=\int_A X_{t_A}dP=\int_A XdP$$

□

3.10. **Rückwärtsmartingale.** Wir haben gesehen, dass ein Martingal genau dann abschließbar ist, wenn es ein maximales Element im Sinne eines L^1-Grenzertes für wachsende Indizes gibt. Analog kann man nach der Existenz eines minimalen Elementes im Sinne eines Grenzwertes für den Fall fragen, dass man in der Indexmenge beliebig weit absteigen kann.

Bevor wir eine Antwort auf diese Frage geben, führen wir der Vollständigkeit halber noch die hier häufig verwendete Sprechweise eines *Rückwärtsmartingals* ein[3].

Definition 133. Eine geordnete Menge $(I,\leq)$ heißt nach unten unbeschränkt, falls für jedes $i\in I$ ein $j\neq i$ existiert mit $j\leq i$.

Definition 134. Ein Martingal $(X_i)_I$ auf einem filtrierten Wahrscheinlichkeitsraum $(\Omega,(\mathcal{J}_i)_{i\in I},\mathcal{J},P)$ mit nach unten unbeschränkter Indexmenge I heißt *Rückwärtsmartingal*

In analoger Weise spricht man dann von Rückwärts-Super- und Rückwärts-Submartingalen. – Von nun an beschränken wir uns auf den Fall $I=\mathbb{Z}_-$. Hier gibt es für Supermartinale das folgende einfache Kriterium für die gleichgradige Integrierbarkeit.

Satz 135. *Für ein Supermartingal $(X_n,\mathcal{J}_n)_{n\in\mathbb{Z}_-}$ sind äquivalent*

[3]Diese Bezeichnung ist genau genommen nicht nur überflüssig sondern auch irreführend, denn eine Umkehrung der Zeit findet dabei nicht statt, trotzdem praktisch.

(1) $\sup_{n\in\mathbb{Z}_-} \mathbb{E}(X_n) < \infty$
(2) $\sup_{n\in\mathbb{Z}_-} \mathbb{E}(|X_n|) < \infty$
(3) $(X_n)_{n\in\mathbb{Z}_-}$ ist glgr. intbar.

Beweis. Wegen $X_n \leq |X_n|$ folgt (1) aus (2). Umgekehrt gilt

$$|X_n| = X_n + 2\left(X_n^-\right) \Longrightarrow \mathbb{E}(|X_n|) = \mathbb{E}(X_n) + 2\mathbb{E}\left(X_n^-\right).$$

Zudem ist

$$Y_n := X_n^- = f(X_n),$$

mit $\mathbb{R} \ni x \to f(x) := -x \vee 0$ ein Submartingal. Somit ist $n \to \mathbb{E}(X_n^-)$ wachsend, also insbesondere

$$\mathbb{E}\left(X_n^-\right) \leq \mathbb{E}\left(X_1^-\right) \leq \mathbb{E}\left(\left|X_1^-\right|\right) < \infty,$$

so dass im Falle von $\mathbb{E}(X_n) < \infty$ schließlich

$$\sup_n \mathbb{E}(|X_n|) \leq \sup_n \mathbb{E}(X_n) + 2\mathbb{E}(|X_1|) < \infty.$$

(3) $\Rightarrow$ (2) folgt unmittelbar, denn eine Familie von gleichgradig integrierbaren Zufallsvariablen ist L^1-beschränkt.
Wir zeigen zuletzt, dass (1) $\Rightarrow$ (2). Da $X.$ ein Supermartingal ist, gilt

$$\sup_{n\in\mathbb{Z}_-} \mathbb{E}(X_n) = \lim_{n\to-\infty} \mathbb{E}(X_n) < \infty.$$

Zu jedem $\varepsilon > 0$ existiert also ein $k_\varepsilon \in \mathbb{Z}_-$ so dass $\sup_{n\in\mathbb{Z}_-} \mathbb{E}(X_n) \leq \varepsilon + \mathbb{E}(X_{k_\varepsilon})$ Hiermit können wir für $n \leq k_\epsilon$ abschätzen

$$\begin{aligned}
\mathbb{E}(|X_n|; |X_n| \geq \alpha) &= \mathbb{E}(X_n; X_n \geq \alpha) - \mathbb{E}(X_n; X_n \leq -\alpha) \\
&= \mathbb{E}(X_n) - \mathbb{E}(X_n; X_n < \alpha) - \mathbb{E}(X_n; X_n \leq -\alpha) \\
&\leq \varepsilon + \mathbb{E}(X_{k_\epsilon}) - \mathbb{E}(X_n \mid X_n < \alpha) - \mathbb{E}(X_n \mid X_n \leq -\alpha) \\
&\leq \varepsilon + \mathbb{E}(X_{n_k}) - \mathbb{E}\left(X_{n_{k_\varepsilon}} \mid X_n < \alpha\right) - \mathbb{E}\left(X_{n_{k_\varepsilon}} \mid X_n \leq -\alpha\right),
\end{aligned}$$

wobei wir im letzten Schritt die Submartingaleigenschaft (und $n \leq k_\epsilon$) ausgenutzt haben, denn $\{X_n < \alpha\} \in \mathcal{J}_n$ und für $A \in \mathcal{J}_n$

$$\int_A X_n dP \geq \int_A \mathbb{E}\left(X_{n_{k_\varepsilon}} \mid \mathcal{J}_n\right) dP = \int_A X_{n_{k_\varepsilon}} dP.$$

Insgesamt gilt für $n \leq k_\epsilon$, dass

$$\mathbb{E}(|X_n| \mid |X_n| \geq \alpha) \leq \varepsilon + \mathbb{E}\left(X_{n_{k_\varepsilon}} \mid |X_n| \geq \alpha\right).$$

Hieraus folgt die gleichgradige Integrierbarkeit von $(X_n)_{n\in\mathbb{Z}}$. □

Korollar 136. *Falls $(X_k)_{k\in\mathbb{Z}_-}$ ein $(\mathcal{J}_k)_{k\in\mathbb{Z}_-}$-Martingal ist, so ist bereits abschließbar in $-\infty$, d.h. $\underline{X} = \lim_{k\to-\infty} X_k$ existiert fast sicher und in L^1 mit*

$$\mathbb{E}(X_j \mid \mathcal{J}_{-\infty}) = \underline{X}, \quad \mathcal{J}_{-\infty} := \bigcap_{k\in\mathbb{Z}_-} \mathcal{J}_k.$$

Beweis. Es ist $k \to \mathbb{E}(X_k) = C$ konstant. Nach dem vorausgehenden Satz ist (X_k) gleichgradig integrierbar, zudem existiert der fast sichere Limes $\underline{X} = \lim_{k\to-\infty} X_k$ wegen der Doob'schen Überkreuzungsungleichung. □

3.11. **Anwendung: Austauschbarkeit.**

Definition 137. Eine endliche Menge $X_1, ..., X_n$ von Zufallsvariablen auf $(\Omega, \mathcal{J}, P)$ mit Werten in $(E, \mathcal{E})$ heißt *austauschbar*, falls

$$\text{Vert}\,(X_1, ..., X_n) = \text{Vert}\left(X_{\sigma(1)}, ..., X_{\sigma(n)}\right)$$

für jede Permutation $\sigma \in S_n$.
Beispiel 138.

(1) Falls $X_i, i = 1, ..., k$ i.i.d., so sind X_i austauschbar.
(2) Gegeben sei eine Urne mit N Kugeln, M davon seien schwarz. Sie werden nacheinander gezogen. Dann sind die Zufallsvariablen $\{X_1, ..., X_n\}$

$$X_j = \begin{cases} 1 & \text{falls j-te Kugel schwarz} \\ 0 & \text{sonst} \end{cases}$$

austauschbar, denn

$$P\left(\overrightarrow{X} = (0, 1, 1, 0, ..., 1)\right) = \frac{1}{\binom{N}{M}}.$$

Für die Definition der Austauschbarkeit einer unendlichen Menge von Zufallsvariablen führen wir die folgenden Sprechweisen ein.

Definition 139. (1) Eine Bijektion $\sigma : \mathbb{N} \to \mathbb{N}$ heißt *endliche Permutation*, falls $\sigma_{[N,\infty[} = \text{id}_{[N,\infty[}$ für ein geeignet groß gewähltes $N \in \mathbb{N}$.
(2) Für eine Folge $(x_k)_{k\in\mathbb{N}}$ mit Werten in E und $\sigma \in S_n$ definieren wir $(x_k^\sigma)_{k\in\mathbb{N}}$ durch $x_k^\sigma = x_{\sigma(k)}$ für $k \leq n$ bzw. $x_k^\sigma = x_k$ für $k > n$.

Definition 140. Ein Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ heißt *austauschbar*, falls $\text{Vert}\,(X_.) = \text{Vert}\,(X_.^\sigma)$ für alle endlichen Permutationen $\sigma : \mathbb{N} \to \mathbb{N}$.
Definition 141.

(1) $f : E^n \to \mathbb{R}$ heißt symmetrisch, falls für alle $X \in E^n$:

$$f(X) = f(X^\sigma)$$

(2) $f : E^{\mathbb{N}} \to \mathbb{R}$ heißt n-symmetrisch, falls (1) gilt für alle gilt für alle $\sigma \in S_n$.

Beispiel 142.

(1) $f\,(X_1, ..., X_n) = \frac{1}{n} \sum_{i=1}^n X_i$
(2) $f\,(X_n) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n X_i$

Definition 143. Es sei $X_n : (\Omega, \mathcal{J}, P) \to (E, \mathcal{E})$, $n \in \mathbb{N}$, ein stochastischer Prozess. Dann heißt

$$\mathcal{E} := \bigcap_{k\geq 0} \mathcal{E}_n \subset \mathcal{J}$$

mit

$$\mathcal{E}_n := \sigma\,(f\,(X_1, ..., X_n) \mid f : E^n \to \mathbb{R} \text{ symmetrisch}) \subset \mathcal{J}$$

die $\sigma-$*Algebra der austauschbaren Ereignisse.*

Bemerkung 144. Für die vom Prozess X_n erzeugte terminale σ-Algebra τ gilt $\tau \subset \mathcal{E}$, denn

$$A \in \sigma\left(X_m, X_{m+1}, ...\right) \subset \mathcal{E}_n, \text{ falls } m \geq n+1.$$

Daher

$$\underbrace{\bigcap_{m\geq 0} \sigma\left(X_l \mid l \geq m\right)}_{=\tau} \subset \mathcal{E}_n \Longrightarrow \tau \subset \bigcap_{n \geq n_0} \mathcal{E}_n = \mathcal{E}.$$

Im Allgemeinen ist ist $\tau \subsetneq \mathcal{E}$, denn falls $\#E \geq 2$ sei $B \subset E$ mit B messbar, $B \neq \varnothing$ und $B = E$. Dann gilt

$$S := \sum_{n=1}^{\infty} 1_B\left(X_n\right) \Longrightarrow S \text{ ist messbar bzgl. } \mathcal{E} \text{ aber nicht bzgl. } \tau.$$

Für $f : E^N \to \mathbb{R}$ und einen stochastischen Prozess $\left(X_k\right)_k$ mit Werten in E und führen wir noch die Schreibweise ein

$$f(X) = f\left(X_1, ..., X_N\right).$$

Satz 145. *Falls* $\left(X_k\right)_k$ *austauschbar und* $f : E^n \to \mathbb{R}$*, dann folgt:*

$$\mathbb{E}\left(f(X) \mid \mathcal{E}_n\right) = \mathbb{E}\left(f\left(X^\sigma\right) \mid \mathcal{E}_n\right), \forall \sigma \in S_n$$

Insbesondere gilt

$$\mathbb{E}\left(f(X) \mid \mathcal{E}_n\right) = \frac{1}{n!} \sum_{\sigma \in S_n} f\left(X^\sigma\right).$$

Beweis. Für $\sigma \in S_n$ und $f : E^n \to \mathbb{R}$ sei $f^\sigma : E^n \to \mathbb{R}$ definiert als

$$f^\sigma(x_1, \cdots, x_n) = f(x_{\sigma_1}, \cdots, x_{\sigma_n}).$$

Sei nun $h : \mathbb{E}^n \to \mathbb{R}$ n-symmetrisch und $\sigma \in S_n$ dann ist $h = h^\sigma$, folglich

$$\begin{aligned}
\mathbb{E}\left(f(X)h(X)\right) &= \mathbb{E}\left(f(X)h^\sigma(X)\right) \\
&= \mathbb{E}\left(f^{\sigma^{-1}}(X^\sigma)h^\sigma(X)\right) \\
&= \mathbb{E}\left(f^{\sigma^{-1}}(X^\sigma)h(X^\sigma)\right) \\
&= \int_{\mathbb{E}^\mathbb{N}} f^{\sigma^{-1}}(z)h(z) \cdot \underbrace{\mathrm{Vert}_{X^\sigma}(dz)}_{=\mathrm{Vert}_{X(dz)}} \\
&= \mathbb{E}\left(f^{\sigma^{-1}}(X)h(X)\right)
\end{aligned}$$

$$\Longrightarrow \mathbb{E}\left(f(X)h(X)\right) = \mathbb{E}\left(f\left(X^{\sigma^{-1}}\right)h(X)\right), \quad \forall \sigma \in S_n$$

Da $\mathcal{E}^n := \sigma\left\{h \mid h\ n\text{-symmetrisch}\right)$, folgt:

$$\mathbb{E}\left(f\left(X^\sigma\right) \mid \mathcal{E}_n\right) = \mathbb{E}\left(f(X) \mid \mathcal{E}_n\right)$$

Die zweite Aussage ergibt sich nun unmittelbar, denn

$$h : E^n \to \mathbb{R}, \quad h(\cdot) := \frac{1}{n!} \sum_{\sigma \in S_n} f^\sigma(\cdot)$$

ist eine symmetrische Funktion, somit $\xi := h(X)$ messbar bzgl. $\mathcal{E}_n$ und nach dem vorausgehenden

$$\mathbb{E}\left(g(X)f(X)\right) = \mathbb{E}\left(g(X)h(X)\right) \quad \forall g\, n\text{-symmetrisch.}$$

□

Um einen Zusammenhang mit Rückwärtsmartingalen herzustellen betrachten wir einen stochastischen Prozess $(X_n)_{n\in\mathbb{N}_0}$ und definieren

$$\mathcal{J}_{-n} := \mathcal{E}_n \subset \mathcal{J}, \quad Y_{-n} := \frac{1}{n}\sum_{k=1}^{n} X_k \text{ für } n \in \mathbb{N}.$$

Y_{-n} ist dann ist messbar bzgl. $\mathcal{J}_{-n}$.

Lemma 146. *Falls $(X_k)_k$ austauschbar ist, so bildet $(Y_k)_{k\in\mathbb{Z}_-}$ ein $(\mathcal{J}_k)_{k\in\mathbb{Z}_-}$-Martingal (und daher auch auch ein $(\mathcal{G}_k := \sigma(Y_l \mid l \leq k))_{k\in\mathbb{Z}_-}$-Martingal).*

Beweis. Wir schreiben $\mathbb{E}\left(Y_{-(n-1)} \mid \mathcal{J}_{-n}\right) = \mathbb{E}\left(\varphi\left(X_1, ..., X_{n-1}\right) \mid \mathcal{E}_n\right)$ mit

$$\begin{aligned} Y_{-(n-1)} &= \frac{1}{n-1}\sum_{k=1}^{n-1} X_k \\ &= \varphi\left(X_1, ..., X_{n-1}\right) \end{aligned}$$

so dass

$$\begin{aligned} \mathbb{E}\left(Y_{-(n-1)} \mid \mathcal{J}_{-n}\right) &= \frac{1}{n!}\sum_{\sigma\in S_n} \varphi\left(X_{\sigma(1)}, ..., X_{\sigma(n-1)}\right) \\ &= \frac{1}{n!}\frac{1}{n-1}\sum_{\sigma\in S_n}\left(X_{\sigma(1)} + ... + X_{\sigma(n-1)}\right) \\ &= \frac{1}{n!}\frac{1}{n-1}\sum_{k=1}^{n}\sum_{\substack{\sigma\in S_n \\ \sigma(n)=k}} \underbrace{\left(X_{\sigma(1)} + ... + X_{\sigma(n)}\right)}_{=\left(\sum_{\substack{l=1 \\ l\neq k}}^{n} X_l\right)(n-1)!} \\ &= \frac{1}{n}\frac{1}{n-1}\sum_{k=1}^{n}\sum_{\substack{l=1 \\ l\neq k}}^{n} X_l \\ &= \frac{1}{n}\sum_{j=1}^{n} X_j \\ &= Y_{-n} \end{aligned}$$

□

Korollar 147 (Starkes Gesetz der großen Zahlen nach Kolmogorov)**.** *Sei $(X_k)_k$ i.i.d. mit $\mathbb{E}\left(|X_1|\right) < \infty$, dann gilt:*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow \mathbb{E}(X) \textit{ fast sicher und in } L^1.$$

Beweis. $Y_{-n} := \frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow Y_{-\infty} := \mathbb{E}\left(Y_{-1} \mid \mathcal{J}_{-\infty}\right) = \mathbb{E}\left(X_1 \mid \mathcal{E}\right)$ fast sicher und in L^1 mit Y_{-n} als Rückwärtsmartingal bzgl. $(\mathcal{J}_{-n})$. Zudem ist $Y_{-\infty}$ messbar bzgl. $\tau \subset \mathcal{E}$. Daher

$$Y_{-\infty} = \mathbb{E}\left(Y_{-\infty} \mid \tau\right) = \mathbb{E}\left(\mathbb{E}\left(X_1 \mid \mathcal{E}\right) \mid \tau\right) = \mathbb{E}\left(X_1 \mid \tau\right) \text{ fast sicher.}$$

Nach dem Kolmogorov'schen 0-1-Gesetz ist τ ist trivial, so dass

$$\mathbb{E}\left(X_1 \mid \tau\right) = \mathbb{E}\left(X_1\right).$$

□

Das Kolmogorov'sche Gesetz der großen Zahlen lässt ohne große Mühe nun noch die folgende Verallgemeinerung zu.

Satz 148. *Es sei $(X_n)_n$ ein austauschbarer Prozess und $\varphi : E^k \to \mathbb{R}$ messbar beschränkt, dann*

$$A_n\left(\varphi\right) = \frac{1}{n!}\sum_{\sigma\in S_n} \varphi\left(X^\sigma\right) \Longrightarrow A_n \to A\left(\varphi\right) \text{ f.s. und in } L^1$$

mit

$$A\left(\varphi\right) = \mathbb{E}\left(\varphi(X) \mid \mathcal{E}\right) = \mathbb{E}\left(\varphi(X) \mid \tau\right).$$

Beweis. Analog wie oben gilt

$$A_n\left(\varphi\right) = \mathbb{E}\left(\varphi(X) \mid \mathcal{E}_n\right) \longrightarrow \mathbb{E}\left(\varphi \mid \mathcal{E}\right)$$

in L^1 und f.s.. Zudem ist $\lim_{n\to\infty} A_n$ ist $\tau-$messbar, denn für $l \in \mathbb{N}$

$$\frac{\#\left\{\sigma \in S_n \mid \{\sigma(1), ..., \sigma(k)\} \subset \{1, ..., l\}\right\}}{n!} \underset{n\to\infty}{\longrightarrow} 0.$$

so dass $\lim_n A_n\left(\varphi\right)$ nicht von $X_1, ..., X_l$ abhängt. □

Auch das Kolmogorov'sche 0-1-Gesetz lässt sich auf austauschbare Ereignisse ausdehnen, wie wir gleich sehen werden. Hierfür benötigen wir das folgende Hilfsresulat.

Satz 149. *Falls $(X_n)_n$ austauschbar und $A \in \mathcal{E}$, so existiert $B \in \tau$ mit*

$$P(A \triangle B) = 0.$$

Beweis. Für $\varphi = 1_A$ gilt

$$\varphi = \lim_n A_n\left(\varphi\right) =: \xi \text{ fast sicher },$$

wobei ξ wie oben gesehen messbar bzgl. τ ist. Zudem folgt aus $\varphi = \xi$ fast sicher, dass $\xi \in \{0, 1\}$ fast sicher. Die Menge $B := \xi^{-1}(1)$ hat daher die gewünschten Eigenschaften. □

Hieraus ergibt sich die angekündigte Verallgemeinerung vom Kolmogorov'schen 0-1-Gesetz.

Korollar 150 (0-1-Gesetz von Hewitt-Savage)**.** *Falls $(X_k)_k$ i.i.d. so gilt für alle $A \in \mathcal{E}$ dass $P(A) = 0$ oder $P(A) = 1$.*

3.12. **Exkurs: Konzentrationsungleichungen.** Das schwache Gesetz der großen Zahlen für i.i.d. Zufallsvariablen $(X_k)_{k\in\mathbb{N}}$ besagt, dass die Streuung der Zufallsgröße $\frac{1}{n}\sum_{i=1}^{n} X_i$ um ihren Erwartungswert $\mu = E(X)$ für hinreichend großes n nahezu Null ist. Die Zufallsvariable $\frac{1}{n}\sum_{i=1}^{n} X_i$ ist also nahezu deterministisch in dem Sinne, dass Abweichungen vom Erwartungswert mit nur verschwindend geringer Wahrscheinlichkeit auftreten. – Nahezu deterministische Zufallsvariablen können auch auf andere Weise denn als Mittelwerte von unabhängigen Zufallsvariablen entstehen, und hier versucht man, die Streuung durch sogenannte Konzentrationsungleichungen möglichst präzise abzuschätzen.

Beispiel 151 (Mittlere Länge eines minimalen Polygonzuges für N Zufallspunkte im Einheitsquadrat)**.** Wir wählen N Punkte im Einheitsquadrat der Ebene

$$X_1, ..., X_N \text{ in } [0,1]^2 \subset \mathbb{R}^2$$

unabhängig voneinander aus. Es bezeichne $L(X_1, ..., X_N)$ die kleinstmögliche Gesamtlänge eines die Punkte $\{X_1, ..., X_N\}$ in beliebiger Reihenfolge verbindenden Polygonzuges (Tour durch die Punkte $\{X_1, \cdots, X_n\}$), dann beschreibt

$$F_N = \frac{1}{N} L\left(\overline{X}_1, ..., \overline{X_N}\right)$$

die mittlere Weglänge (pro Punkt) des Punktensembles $\{X_1, \ldots, X_N\}$. Wir zeigen am Ende dieses Abschnittes, dass F_N nahezu deterministisch ist, denn es gilt die Abschätzung

$$P\left(|F_N - \mathbb{E}(F_N)| > \varepsilon\right) \le e^{-\varepsilon N}.$$

Insbesondere ist die Wahrscheinlichkeit einer Abweichungen der Zuflalsvariable F_N von ihrem Mittelwert exponentiell unwahrscheinlich.

Der Beweis hiervon ergibt sich am Ende des Abschnitts als einfache Anwendung der McDiarmid-Konzentrationsungleichung. Ausgangspunkt ist die folgende Abschätzung.

Lemma 152 (Hoeffding-Lemma)**.** *Sei X eine reelle Zufallsvariable und $\mathbb{E}(X) = 0, X \in [a,b]$ fast sicher. Dann gilt*

$$\mathbb{E}\left(e^{sX}\right) \le e^{\frac{s^2(b-a)^2}{8}}.$$

Beweis. $X \to e^{sX}$ ist konvex mit $s \in \mathbb{R}$. Dann:

$$X = \alpha b + (1-\alpha)a,\ \alpha = \frac{X-a}{b-a}$$

$$\begin{aligned} \Longrightarrow e^{sX} &\le \alpha e^{sb} + (1-\alpha)e^{sa} \\ &= \frac{X-a}{b-a}e^{sb} + \frac{b-X}{b-a}e^{sa} \end{aligned}$$

$$\begin{aligned} \Longrightarrow \mathbb{E}\left(e^{sX}\right) &\le \frac{\mathbb{E}(X)-a}{b-a}e^{sb} + \frac{b-\mathbb{E}(X)}{b-a}e^{sa} \\ &= \frac{-a}{b-a}e^{sb} + \frac{b}{b-a}e^{sa} \\ &\overset{!}{\le} e^{\frac{s^2(b-a)^2}{8}} \end{aligned}$$

Zum Beweis der letzten Ungleichung schreiben wir

$$\frac{-a}{b-a}e^{sb} + \frac{b}{b-a}e^{sa} := e^{L(h)}$$

wobei $h = s(b-a)$ und $\eta := \frac{-a}{b-a}$, so dass

$$L(h) = -h\eta + \ln(1 - \eta + \eta e^h).$$

Man rechnet nun nach, dass $L(0) = 0, L'(0) = 0, L''(h) \leq \frac{1}{4}$, woraus $L(h) \leq \frac{1}{8}h^2$ folgt. $\square$

Bemerkung 153. Für das nächste Resultat benötigen wir das Hoeffding-Lemma in der Version für die bedingte Erwartung, wonach für eine Zufallsvariable $X \in [a,b]$ f.s. und $E(X|\mathcal{G}) = 0$ für $s \in \mathbb{R}$ gilt, dass

$$\mathbb{E}(e^{sX}|\mathcal{G}) \leq e^{\frac{s^2(b-a)^2}{8}}.$$

Der Beweis ist identisch zum vorausgehenden.

Satz 154 (Azuma-Hoeffding)**.** *Es sei* $(X_k)_{k=0,\ldots,N}$ *ein* $(\mathcal{G}_k)_{k=0,\ldots,N}$*-Martingal mit Start in* $X_0 = 0$ *und*

$$\forall k \geq 1 \quad |X_k - X_{k-1}| \leq c_k \textit{ fast sicher.}$$

Dann gilt für $\lambda \in \mathbb{R}$, $n \in \{1, \cdots, N\}$

$$\mathbb{E}\left(e^{\lambda X_n}\right) \leq e^{\frac{1}{2}\lambda^2 \sum_{k=1}^n c_k^2}$$

Beweis. Durch Induktion. Für $k = 1$ (mit $X_0 = 0$) ist $X_1 \in (-c_1, c_1)$ fast sicher und $\mathbb{E}(X_1) = 0$. Die Behauptung folgt mit Hoeffding. Für den Induktionsschritt schreiben wir

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda X_n}\right) &= \mathbb{E}\left(e^{\lambda X_{n-1}} e^{\lambda(X_n - X_{n-1})}\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\ldots\right) \mid \mathcal{G}_{n-1}\right) \\
&= \mathbb{E}\left(e^{\lambda X_{n-1}}\mathbb{E}\left(e^{\lambda\overbrace{(X_n - X_{n-1})}^{Z_n}} \mid \mathcal{G}_{n-1}\right)\right).
\end{aligned}$$

Für Z_n gilt:

$$\begin{array}{ll}
1) \quad \mathbb{E}\left(Z_n \mid \mathcal{G}_{n-1}\right) & = 0 \\
2) \quad Z_n \in (-c_n, c_n) & \text{fast sicher.}
\end{array}$$

Mit dem Hoeffding Lemma (in der Version für bedingte Erwartungen) gilt

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda Z_n} \mid \mathcal{G}_{n-1}\right) &\leq e^{\frac{1}{8}\lambda^2(c_n - (-c_n))^2} \\
&= e^{\frac{1}{2}\lambda^2 c_n^2},
\end{aligned}$$

und somit

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda X_n}\right) &\leq \mathbb{E}\left(e^{\lambda X_{n-1}} e^{\frac{\lambda^2}{2}c_n^2}\right) \\
&= e^{\lambda^2 \frac{c_n^2}{2}}\mathbb{E}\left(e^{\lambda X_{n-1}}\right) \\
&\leq e^{\frac{\lambda}{2}\sum_{j=1}^n C_j^2}.
\end{aligned}$$

$\square$

Korollar 155 (Martingal-Konzentrationsungleichung)**.** *In der Situation wie oben gilt für $\epsilon > 0$*

$$P(X_n > \varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2\sum_{j=1}^n c_j^2}\right).$$

Beweis. Wir setzen $u := \sum_{j=1}^n c_j^2$, dann schließen wir der exponentiellen Tschebbyschev-Ungleichung und Azuma-Hoeffding, dass

$$\begin{aligned} P(X_n > \varepsilon) &\underset{\lambda>0}{=} P\left(e^{\lambda X_n} > e^{(\lambda\varepsilon)}\right) \\ &\leq e^{-\lambda\varepsilon}\mathbb{E}\left(e^{\lambda X_n}\right) \\ &\leq e^{-\lambda\varepsilon}e^{\frac{\lambda^2 u}{2}} \\ &= e^{\frac{\lambda^2 u}{2}-\lambda\varepsilon}. \end{aligned}$$

Die Minimalstelle bzgl. $\lambda > 0$ auf der rechten Seite, bzw. gleichbedeutend die Minimalstelle von

$$\lambda \longrightarrow \frac{\lambda^2 u}{2} - \lambda\varepsilon,$$

ist $\lambda^* := \frac{\varepsilon}{u}$, welches eingesetzt ergibt, dass

$$P(X_n > \varepsilon) \leq e^{-\frac{\varepsilon^2}{2u}}.$$

□

Korollar 156 (McDiarmid-Ungleichung)**.** *Seien $X_1, ..., X_n$ unabh. mit Werten in E. Sei $F : E \times ... \times E = E^n \to \mathbb{R}$, s.d. für alle $x_1, .., x_n, y \in E^{n+1}$:*

$$|F(x_1, ..., x_{l-1}, y, x_{l+1}, ..., x_n) - F(x_1, ..., x_n)| \leq c_l$$

Dann gilt für die Zufallsvariable $\Phi := F(X_1, ..., X_n)$

$$P(|\Phi - \mathbb{E}(\Phi)| > \varepsilon) \leq 2e^{-\frac{\varepsilon^2}{2\sum_{j=1}^n c_j^2}}$$

Beweis. OBdA ist $\mathbb{E}(\Phi) = 0$, denn andernfalls ersetzen wir F durch $\tilde{F} := F - c$ mit $c = \mathbb{E}(\Phi)$. Sei $\mathcal{G}_j := \sigma(X_1, ..., X_j)$. Sei $M_j := \mathbb{E}(\Phi \mid \mathcal{G}_j)$ und $M_0 := 0$. Dann gilt

$$\begin{array}{ll} 1) & (M_j)_{j=1,...,n} \text{ ist } \mathcal{G}\text{-Martingal} \\ 2) & M_N = \mathbb{E}(\Phi \mid \sigma(X_1, ..., X_N)) = F(X_1, ..., X_N) = \Phi \\ 3) & |M_j - M_{j-1}| \leq c_j, \text{ f.s.} \end{array}$$

Um die Eigenschaft 3) einzusehen, verwenden wir eine einfache Hilfsausage, die wir zum Zwecke der Übersichtlichkeit gesonderst formulieren wollen.

Lemma 157. *Es seien X, Y unabhängige Zufallsvariablen und $Z := F(X,Y)$ eine messbare Funktion von X, Y, dann gilt $E(Z|\sigma(X))(\omega) = \mathbb{E}_{\tilde{Y}}[F(X(\omega), \tilde{Y})]$ für P-fast alle ω, wobei $\tilde{Y} \simeq Y$ eine unabhängige wie Y verteilte Zufallsvariable ist.*

Die Aussage ist trivial im Fall, dass $F(x,y) = 1_A(x) \cdot 1_B(y)$ ist trivial; der allgemeine Fall folgt mit einem monotone Klasse-Argument.

Sei nun $Z_1, \cdots, Z_N$ eine unabhängige Familie von Zufallsvariablen mit $Z_i \simeq X_i$ für $i = 1, \cdots, N$. Dann gilt gemäß dem vorausgehenden Lemma

$$\begin{aligned}\mathbb{E}(\Phi|\mathcal{G}_j) &= \mathbb{E}\left(F(X_1, \cdots, X_N) \mid \sigma\left(X_1, ..., X_j\right)\right) \\ &= \mathbb{E}_Z\left(F\left(X_1, ..., X_j, Z_{j+1}, ..., Z_N\right)\right) \\ &= \int_E ... \int_E F\left(X_1(\omega), ..., X_j(\omega), z_j, ..., z_N\right) \mathrm{Vert}_{X_{j+1}}(dz_{j+1}) ... \mathrm{Vert}_{X_N}(dz_N).\end{aligned}$$

Folglich

$$\begin{aligned}M_j - M_{j-1} &= \mathbb{E}_Z\left(F\left(X_1, ..., X_j, Z_{j+1}, ..., Z_N\right)\right) - \mathbb{E}_Z\left(F\left(X_1, ..., X_{j-1}, Z_j, ...\right)\right) \\ &= \mathbb{E}\left(F\left(X_1, ..., X_j, Z_{j+1}, ...\right) - F\left(X_1, ..., X_{j-1}, Z_j, ...\right)\right) \\ &\leq \mathbb{E}\left(\underbrace{|F(...) - F(...)|}_{\leq c_j}\right) \\ &\leq c_j\end{aligned}$$

Der Satz ergibt sich nun aus der Azuma-Hoeffding-Ungleichung. □

Aus der McDiarmid-Ungleichung folgt nun direkt die eingangs gemachte Aussage über die Konzentration der mittleren Tourlänge für N zufällig gewählte Punkte im Einheitsquadrat, denn für $F(x_1, \cdots, x_N) = \frac{1}{N} L(x_1, \ldots, x_n)$ gilt offenbar, dass

$$\left|F\left(x_1, ..., x_{l-1}, y, x_{l+1}, ..., x_n\right) - F\left(x_1, ..., x_n\right)\right| \leq \frac{2\sqrt{2}}{N}.$$

4. Brown'sche Bewegungen

4.1. **Vorbemerkung.** In einer Veröffentlichung aus dem Jahre 1828 berichtete der Botaniker Robert Brown von einer unter dem Mikroskop beobachteten zufälligen Zitterbewegung von Pollenkörnern in einer Konservierungsflüssigkeit. Es dauerte noch weitere 70 Jahre bis die ersten mathematischen Modelle hierfür von Louis Bachelier und Albert Einstein sowie später von Norbert Wiener entwickelt wurden. Heutzutage werden diese Modelle als Brown'sche (Molekular-)Bewegungen bezeichnet. Wir wir später sehen werden, handelt es sich hierbei um zeitabhängige Erweiterungen der Normalverteilung. Brown'sche Bewegungen treten in unzähligen Anwendungen angrenzender Wissenschaften auf, etwa zur Beschreibung von Molekülen oder der Ausbreitung von Wärme, oder als Modell für zufällige Kursschwankungen von Wertpapieren in der Finanzmathematik.

Wir beginnen mit einer axiomatischen Definition, wobei wir uns der Einfachheit halber auf den Fall einer reellwertigen d.h eindimensionalen Brown'schen Bewegung beschränken. – Brown beobachtete dagegen eine Zitterbewegung eines Pollens in zwei bzw. drei Dimensionen. Wir stellen uns hierfür vor, dass im räumlichen Fall die Bewegung durch unabhängige eindimensionale Brown'sche Bewegungen in den jeweiligen Koordinatenrichtungen entstehen, weshalb es zunächst ausreicht, die Bewegung in den einzelnen Koordinaten sinnvoll beschreiben zu können.

In moderner Terminologie suchen wir einen stochastischen Prozess $t \mapsto X_t(\omega) \in \mathbb{R}, \omega \in \Omega$, in stetiger Zeitparametrisierung, welcher die folgenden Kernforderungen erfüllt.

(1) Die Zuwächse des Prozesses zwischen zwei beliebigen Beobachtungszeitpunkten sind zentriert normalverteilt, d.h.

$$X_{t+h} - X_t \cong \nu_{0,h}$$

(Standard-Normalverteilung mit Erwartungswert 0 und Varianz $v = h$.).

(2) Die Zuwächse von $X.$ sind stochastisch unabhängig, d.h. für eine beliebige Auswahl an Beobachtungszeitpunkten

$$0 \leq t_1 \leq ... \leq t_N$$

sind die Zufallsvariablen $\{Z_1, \cdots, Z_N\}$

$$Z_1 := X_{t_1} - X_0, ..., Z_N := X_{t_N} - X_{t_{N-1}}$$

stochastisch unabhängig.

(3) Die *Trajektorien* $t \mapsto X_t = X_t(\omega)$ sind für P-fast jedes ω stetig.

Definition 158. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $X_t : (\Omega, \mathcal{F}) \to \mathbb{R}, t \in \mathbb{R}_{\geq 0}$ ein stochastischer Prozess. Dann heißt der Prozess $(X_t)_{t\geq 0}$ eine *Brown'sche Bewegung*, falls er die Bedingungen (1) – (3) erfüllt. Falls zusätzlich $X_0 = 0$ fast sicher, so heißt (X_t) eine *standard Brown'sche Bewegung*.

Bemerkung 159. (1) Die weiteren Eigenschaften des Wahrscheinlichkeitsraumes $(\Omega, \mathcal{F}, P)$ spielen keine Rolle bei der Frage, ob der hierauf definierte Prozess $(X_t)_{t\geq 0}$ eine Brown'sche Bewegung ist.

(2) Mit der obigen Definition wird der Begriff 'Brownsche Bewegung' als Bezeichnung einer *Klasse von stochastischen Prozessen* eingeführt.

Zunächst wollen wir klären, dass wir nicht von der leeren Menge sprechen.

Satz 160. *Die Klasse der Brown'schen Bewegungen ist nicht leer.*

Zum Beweis geben wir im Folgenden eine explizite Konstruktion einer Brownschen Bewegung an, die jedoch einige Vorbereitungen hinsichtlich mehrdimensionaler Gaußmaße erfordert.

4.2. **Erinnerung: Gaußmaße auf dem Euklidischen Raum.** Wir erinnern zunächst an die Normalverteilung auf der reellen Achse.

Definition 161. Das Maß $\nu_{m,v}(dx) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-m)^2}{2v}} dx$ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißt *Normalverteilung* auf $\mathbb{R}$ mit Erwartungswert m und Varianz $v \geq 0$.

Aus der Warhscheinlichkeitsteorie I wissen wir, dass $\int_{\mathbb{R}} \nu_{m,v}(dx) = 1$, $\int_{\mathbb{R}} x\nu_{m,v}(dx) = m$ und $\int_{\mathbb{R}}(x-m)^2\nu_{m,v}(dx) = v$. – Im Fall $v = 0$ setzen wir $\nu_{m,v} = \delta_m(dx)$ (Dirac-Punktmaß im Punkt $m \in \mathbb{R}$.
Wir führen jetzt noch die *Laplace-Transformation* als sehr nützliches Hilfsmittel zur Untersuchung und Identifikation von Maßen auf $\mathbb{R}$ ein. Es handelt sich dabei um einen engen Verwandten der charakteristischen Funktion (Fourier-Transformation), die wir bereits in der Wahrscheinhlichkeitstheorie I kennengelernt haben.

Definition 162. Es sei ν ein Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, dann heißt die Funktion

$$\Lambda_\nu : \mathbb{R} \to [0, \infty], \quad \lambda_\nu(t) := \int_{\mathbb{R}} e^{tx} \nu(dx).$$

die *Laplace-Transformierte von* ν. Falls X eine reelle Zufallsvariable mit Verteilunng ν ist, so gilt

$$\lambda_\nu(t) = \mathbb{E}(e^{tX}) =: \Lambda_X(t),$$

und wir sprechen von der *Laplace-Transformierten bzw. Momenten-erzeugenden Funktion von* X.

Formal gilt $\Lambda(it) = \int_{\mathbb{R}} e^{itx} \nu(dx)$, was der der charakteristischen Funktion entspricht, d.h. die Laplace-Transformation und die die charakteristische Funktion von ν gehen aus der Einschränkung der verallgemeinerten charakteristischen Funktion

$$\mathbb{C} \ni z \to \hat{\nu}(z) := \int_{\mathbb{R}} e^{xz} \nu(dx)$$

auf die reelle bzw. imaginäre Achse in $\mathbb{C}$ hervor. – Die Menge $D(\Lambda_\nu) = \{t \in \mathbb{R} \mid \Lambda_\nu(t) < \infty\} = I \subset \mathbb{R}$ ist der *Definitionsbereich* von λ_ν. Im Innern I° von I ist Λ_ν reell analytisch, d.h. als absolut konvergente Potenzreihe darstellbar. In der einer Umgebung von $I^\circ \subset \mathbb{C}$ ist $\hat{\nu}$ somit holomorph, also komplex analytisch. Wegen $o \in I^\circ$ bestimmt die Laplace-Transformierte von ν die charakteristische Funktion von ν eindeutig und umgekehrt. Zusammen mit dem Eindeutigkeitssatz von Levy für charakteristische Funktionen ergibt sich schließlich die folgende Aussage.

Satz 163. *Jedes Borel-Maß auf* $\mathbb{R}$ *ist durch seine Laplace-Transformierte eindeutig bestimmt*

Beispiel 164 (Exponentialverteilung zum Parameter $\lambda > 0$)**.** Es sei $\nu(dx) = \lambda e^{-\lambda x} 1_{\{x \geq 0\}}(dx)$. Dann berechnen wir

$$\lambda_\nu(t) = \lambda \int_{\mathbb{R}} e^{tx} e^{-\lambda x} dx < \infty \iff t \geq \lambda,$$

d.h.

$$\Lambda_\nu(t) = \begin{cases} \infty & \text{falls } t \geq \lambda \\ \frac{\lambda}{t-\lambda} & \text{sonst.} \end{cases}$$

Lemma 165. $\Delta_{m,v}(t) = \mathbb{E}_{m,v}\left(e^{tx}\right)$ *mit* $X \cong U_{m,v}, X \cong \sqrt{v} Z + m$, *wobei* $Z \cong \nu_{0,1}$.

Beweis.

$$
\begin{aligned}
\mathbb{E}_{0,1}\left(e^{t\sqrt{v}Z+m}\right) &= e^{tm}\mathbb{E}_{0,1}\left(e^{t\sqrt{v}Z}\right)\\
&= e^{tm}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{t\sqrt{v}x}e^{-\frac{x^2}{2}}dx\\
&= e^{tm}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{\frac{2t\sqrt{v}x}{2}-\frac{x^2}{2}-\frac{(t\sqrt{v})^2}{2}+\frac{(t\sqrt{v})^2}{2}}dx\\
&= e^{tm}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{(x-t\sqrt{v})^2}{2}}dxe^{\frac{t^2v}{2}}\\
&= e^{tm}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}dxe^{\frac{t^2v}{2}}\\
&= e^{tm+\frac{vt^2}{2}}
\end{aligned}
$$

□

Bemerkung 166. Formal: $\Delta_{m,v}(it) = e^{itm-\frac{vt^2}{2}}$.

Beispiel 167. Falls $X \cong \nu_{m_1,v_1}, Y \cong \nu_{m_2,v_2}$, X und Y sind stochastisch unabh.. Dann: $X+Y \cong \nu_{m_1+m_2,v_1+v_2}$

Beweis. $Z := X+Y$. Dann:

$$
\begin{aligned}
\mathbb{E}\left(e^{tz}\right) &= \mathbb{E}\left(e^{t(x+y)}\right)\\
&= \mathbb{E}\left(e^{tx}\right)\mathbb{E}\left(e^{ty}\right)\\
&= e^{t(m_1+m_2)+\frac{t^2}{2}(v_1+v_2)}
\end{aligned}
$$

□

Definition 168. $b \in \mathbb{R}^d, A \in \mathbb{R}^{d\times d}$ symm., strikt pos. definit. Dann:

$$
\nu_{b,A}(dx) = \frac{1}{\sqrt{2\pi}^d\sqrt{\det A}}e^{-\frac{\langle x-b,A^{-1}(x-b)\rangle_{\mathbb{R}^d}}{2}}dx_1,...,dx_d
$$

heißt multivariate Gaußverteilung in $\mathbb{R}^d$ mit EW $b \in \mathbb{R}^d$, Kovarianzmatrix $A \in \mathbb{R}^{d\times d}$.

Definition 169. $t \in \mathbb{R}^d$. Dann: $\Lambda(t) = \mathbb{E}_{\nu_{b,A}}\left(e^{\langle t,x\rangle}\right) = \frac{1}{\sqrt{2\pi}^d\sqrt{\det A}}\int_{\mathbb{R}^d} e^{\langle t,x\rangle}e^{-\frac{\langle (x-b)A^{-1}(x-b)\rangle}{2}}dx$

Behauptung: $\Lambda(t) = e^{\langle b,t\rangle+\frac{1}{2}\langle t,At\rangle}$. Schreibweise $||t||_A^2 := \langle t, At\rangle$ neue Norm auf $\mathbb{R}^d$. Begründung: $b=0, t=0 \Rightarrow \Lambda(0) = \mathbb{E}(e^0) = 1$. Sei B so gewählt, dass

$B^T B = A$ mit B als Martix bestehend aus der ONB zu $\langle,\rangle_A$ (Cholesky)

$$\begin{aligned}
\Longrightarrow \Lambda(0) &= \frac{1}{\sqrt{2\pi}^d\sqrt{\det A}}\int_{\mathbb{R}^d} e^{\frac{\langle x,A^{-1}x\rangle}{2}}dx \\
&\overset{x=B^Ty}{=} \frac{\det B}{\sqrt{2\pi}^d\sqrt{\det A}}\int_{\mathbb{R}^d} e^{-\langle B^Ty,A^{-1}B^Ty\rangle}dy \\
&= \frac{1}{\sqrt{2\pi}^d}\int_{\mathbb{R}^d} e^{-\langle \underbrace{BB^T}_{=A^T} y,A^{-1}y\rangle}dy \\
&= \frac{1}{\sqrt{2\pi}^d}\int_{\mathbb{R}^d} e^{-\langle y,AA^{-1}y\rangle}dy \\
&= \frac{1}{\sqrt{2\pi}^d}\int_{\mathbb{R}^d} e^{-||y||^2}dy \\
&= \frac{1}{\sqrt{2\pi}^d}\int_{\mathbb{R}^d} e^{-\sum_{i=0}^d y_i^2}dy_1,...,dy_d \\
&= \frac{1}{\sqrt{2\pi}^d}\int_{\mathbb{R}^d} e^{-\frac{y_1^2}{2}}dy_1,...,\int_{\mathbb{R}^d} e^{-\frac{y_d^2}{2}}dy_d \\
&= 1
\end{aligned}$$

$\Rightarrow \nu_{b,A}\left(\mathbb{R}^d\right) = 1$. Allgemeiner Fall: $\Lambda(t) = e^{\langle t,b\rangle + \frac{1}{2}\langle t,At\rangle}$ durch quadratische Ergänzung im Integral:

$$\begin{aligned}
\mathbb{E}\left(e^{\langle t,x\rangle}\right) &= \frac{1}{\sqrt{2\pi}^d\sqrt{\det A}}\int_{\mathbb{R}^d} e^{\langle t,x\rangle}e^{-\frac{1}{2}\langle (x-b),A^{-1}(x-b)\rangle}dx \\
&= \frac{1}{\sqrt{2\pi}^d\sqrt{\det A}}e^{\langle t,b\rangle}\int_{\mathbb{R}^d} e^{\langle t,x-b\rangle}e^{-\frac{1}{2}\langle (x-b),A^{-1}(x-b)\rangle}dx \\
&= \sim\int_{\mathbb{R}^d} e^{\langle t,x-b\rangle}e^{-\frac{1}{2}\langle (x-b),A^{-1}(x-b)\rangle}dx \\
&= \sim\int_{\mathbb{R}^d} e^{\langle t,x\rangle-\frac{1}{2}\langle x,A^{-1}x\rangle}dx \\
&= \sim\int_{\mathbb{R}^d} e^{\langle A^Tt,A^{-1}x\rangle-\frac{1}{2}\langle x,A^{-1}x\rangle}dx \\
&= \sim\int_{\mathbb{R}^d} e^{\langle A^Tt,x\rangle_{A^{-1}}-\frac{1}{2}\langle x,x\rangle_{A^{-1}}} \\
&= \sim\int_{\mathbb{R}^d} e^{\langle A^Tt,x\rangle_{A^{-1}}-\frac{1}{2}\langle x,x\rangle_{A^{-1}}-\frac{1}{2}\langle A^Tt,A^Tt\rangle_{A^{-1}}}dx - e^{\frac{1}{2}\langle A^Tx,A^Tx\rangle_{A^{-1}}} \\
&= \sim e^{\frac{1}{2}\langle x,Ax\rangle}\int_{\mathbb{R}^d} e^{-\frac{1}{2}\langle x-A^Tt,A^{-1}(x-A^Tt)\rangle}dx \\
&= \sim e^{\frac{1}{2}\langle x,Ax\rangle}\int_{\mathbb{R}^d} e^{-\frac{1}{2}\langle x,A^{-1}x\rangle}dx \\
&= e^{\langle t,b\rangle}e^{\frac{1}{2}\langle x,Ax\rangle}
\end{aligned}$$

Korollar 170. *Sei ν ein WMaß auf $\mathbb{R}^d$. Dann ist $\nu = \nu_{b,A}$, gdw. für alle $z \in \mathbb{R}^d$ die ZV $X(\omega) := \langle x, Z\rangle_{\mathbb{R}^d}$ $X : \mathbb{R}^d \to \mathbb{R}$ Standard-normalverteilt ist mit $Z \cong \nu_{\langle b,Z\rangle,\langle Z,AZ\rangle}$.*

$\nu \in \mathcal{P}\left(\mathbb{R}^d\right)$, d.h. ν ist W-Maß auf $\mathbb{R}^d$.

(1) Interpretation: ν entspricht der Wkeit, einen Punkt $p \in \mathbb{R}^d$ zufällig zu wählen.
(2) Interpretation: $(\Omega, \mathcal{J}, P) = \left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right), \nu\right)$ als Wkeitsraum.

1. ist enthalten in 2.: Sei $X : \Omega \to \mathbb{R}^d$ eine ZV mit Werten in $\mathbb{R}^d$ mit $X(\omega) = \omega$. Dann gilt insb.: ν als W-Maß auf $\mathbb{R}^d$ entspricht der Verteilung eines d-dimensionalen Zufallsvektors. Dann ist:

$$\Lambda(t) = \mathbb{E}_\nu\left(e^{\langle t,X\rangle}\right) = \int_{\mathbb{R}^d} e^{\langle t,x\rangle}\nu(dx),\ t \in \mathbb{R}^d$$

Falls $\nu = \nu_{b,A}$ (d.h. Gaußmaß mit Eigenwerten $b \in \mathbb{R}^d$, Kovarianzmatrix $A \in \mathbb{R}^{d\times d}$ pos. symm.)

$$\Lambda(t) = e^{\langle t,b\rangle + \frac{1}{2}\langle t,At\rangle}$$

bestimmt dann das eind. d-dimensionale Gaußmaß.

Korollar 171. *X als d-dimensionale ZV $X : \Omega \to \mathbb{R}^d$ ist $\nu_{b,A}-$verteilt, gdw. jede Projektion*

$$\pi_t(X) := \langle t, X\rangle \in \mathbb{R}$$

ist reell normalverteilt mit

$$\begin{aligned} \mathbb{E}\left(\pi_T(X)\right) &= \langle t, b\rangle \\ \mathbb{V}\left(\pi_t(X)\right) &= \langle t, At\rangle \end{aligned}$$

Beweis.

$$\begin{aligned} \mathbb{E}\left(e^{\lambda\pi_t(X)}\right) &= \mathbb{E}\left(e^{\langle\lambda t,b\rangle}\right) \\ &= e^{\langle\lambda t,b\rangle+\frac{1}{2}\langle\lambda t,A\lambda t\rangle} \\ &= e^{\lambda\langle t,b\rangle+\frac{\lambda^2}{2}\langle t,At\rangle} \end{aligned}$$

Dann gilt äquivalent: $\pi_t(X) \cong \nu_{\langle t,b\rangle,\langle t,At\rangle}$. □

Korollar 172.

(1) Sei $N = \overrightarrow{N} = \begin{pmatrix} n_1 \\ \vdots \\ n_d \end{pmatrix}$ mit $n_1, ..., n_d$ i.i.d, $n_i \cong \nu_a$. Falls $B \in \mathbb{R}^{d\times d}, b \in \mathbb{R}^d$, dann folgt:

$$X_i := BN + b \cong \nu_{b,BB^T}$$

(2) Falls $X \cong \nu_{b,A}, X = (X_1, ..., X_d)^T$, so sind $\{X_1, ..., X_d\}$ unabh., gdw.

$$A = \operatorname{diag}\left(a_1, ..., a_d\right)$$

(3) Falls $X \cong \nu_{b,A}$, so gilt:

$$\begin{aligned} 1)\, \mathbb{E}\left(X_i\right) &= b_i \\ 2)\, \operatorname{Kov}\left(X_i, X_j\right) &= A_{ij} \end{aligned}$$

Beweis. 3.1) Wähle $t := e_i$ in $\Lambda(\varepsilon t)$ mit $\varepsilon > 0$. Dann:

$$\mathbb{E}\left(e^{\varepsilon\langle t,X\rangle}\right) = e^{\varepsilon\langle t,b\rangle+\frac{1}{2}\varepsilon^2\langle t,At\rangle}$$

Differentiation nach ε an Stelle $\varepsilon = 0$ und Benutzung von monotoner Konvergenz ergibt:

$$\mathbb{E}\left(\langle t, X\rangle e^{\varepsilon\langle t,X\rangle}\right)\Big|_{\varepsilon=0} = \left(\langle t, b\rangle + \varepsilon\langle t, At\rangle\right) e^{\varepsilon\langle t,b\rangle+\frac{\varepsilon^2}{2}\langle t,At\rangle}\Big|_{\varepsilon=0}$$

Dann:

$$\mathbb{E}\left(\langle t, X\rangle\right) = \langle t, b\rangle$$

3.2) OBdA sei $b = 0$. Dann zeige $\mathbb{E}\left(X_i X_j\right) = A_{ij}$. Wähle

$$t := \varepsilon e_i + \delta e_j,\ \varepsilon, \delta > 0$$

$$\begin{aligned}
\mathbb{E}\left(e^{\langle t,X\rangle}\right) &= \mathbb{E}\left(e^{\varepsilon X_i} e^{\delta X_j}\right) \\
&= e^{\varepsilon b_i + \delta b_j} e^{\frac{1}{2}\langle t_0, A t_0\rangle}
\end{aligned}$$

mit $t_0 = \begin{pmatrix} 0 \\ \varepsilon \\ \vdots \\ \delta \\ \vdots \\ 0 \end{pmatrix}$. Differenzieren nach ε und δ in $\varepsilon = 0 = \delta$ liefert das gewünschte Ergebnis. 1. Sei $s = B^T t$

$$\begin{aligned}
\mathbb{E}\left(e^{\langle t, BN+b\rangle}\right) &= e^{\langle t,b\rangle}\mathbb{E}\left(e^{\langle s,N\rangle}\right) \\
&= e^{\langle t,b\rangle}\mathbb{E}\left(e^{\sum_{i=1}^d s_i n_i}\right) \\
&= e^{\langle t,b\rangle}\prod_{i=1}^d \mathbb{E}\left(e^{s_i n_i}\right) \\
&= e^{\langle t,b\rangle}\prod_{i=1}^d e^{s_i^2 \frac{1}{2}} \\
&= e^{\langle t,b\rangle} e^{\frac{1}{2}\sum s_i^2} \\
&= e^{\langle t,b\rangle} e^{\frac{1}{2}||s||^2} \\
&= e^{\langle t,b\rangle} e^{\frac{1}{2}\langle s,s\rangle} \\
&= e^{\langle t,b\rangle} e^{\frac{1}{2}\langle B^T t, B^T t\rangle} \\
&= e^{\langle t,b\rangle} e^{\frac{1}{2}\langle t, BB^T t\rangle} \\
&= e^{\langle t,b\rangle} e^{\frac{1}{2}\langle t, At\rangle}
\end{aligned}$$

2. Für allgemeines $A \in \mathbb{R}^{d\times d}$ pos. symm., $S \in \mathbb{R}^d$ und oBdA $b = 0$ ist

$$\nu_{0,A}(dx) = \frac{1}{\sqrt{2\pi}^d \sqrt{\det A}} e^{-\langle x, A^{-1}x\rangle} dx$$

die gemeinsame Verteilung von $(X_1, ..., X_d)$, d.h. $(X_1, ..., X_d)$ sind stochastisch unabh., gdw. $\mathrm{Vert}\left(\overrightarrow{X}\right) = \otimes_{i=1}^n \mathrm{Vert}_{X_i}$, d.h. $(X_1, ..., X_d)$ sind unabh., gdw. $\varphi\left(X_1, ..., X_d\right) = \varphi\left(X_1\right) \cdot ... \cdot \varphi\left(X_d\right)$ mit gemeinsamen Fnk. $\varphi_i : \mathbb{R} \to \mathbb{R}$.

$$\ln\left(\varphi\left(X_1, ..., X_d\right)\right) = \prod_{i=1}^d \ln\left(\varphi\left(X_i\right)\right)$$

Die linke Seite ist gleich $\langle x, A^{-1}x\rangle$, was einem Polynom zweiter Ordnung entspricht. Die Gleichung gilt gdw., falls A^{-1} bzw. A diagonal. $\square$

Später brauchen wir zwei Aussagen:

(1) Linearkombinationen von normalverteilten ZV sind wieder normalverteilt.

(2) Zwei normalverteilte ZV $X_1, ..., X_d$ auf demselben WRaum $(\Omega, \mathcal{J}, P)$ sind stochastisch unabh., also die Kov. sind paarweise identisch Null.

4.3. **Konstruktion der Brownschen Bewegung nach Levy.**

(1) Zuwächse sind unabhängig.
(2) Zuwächse sind reell normalverteilt (zentriert, Varianz=Zeitspanne)
(3) $t \longrightarrow B_t(\omega)$ stetig für fast alle $\omega \in \Omega$.

Satz 173. *Die BB ex. als widerspruchsfreies mathematisches Objekt.*

Beweis. Konstruiere zunächst $(B_t)_{t\in[0,1]}$: Hierfür konstruiere $B_{|D}$ mit $D = \bigcup_{n\geq 0} D_n$ mit:

$$D_n := \left\{ \frac{k}{2^n}; k = 0, ..., 2^{n-1} \right\}$$

Konstruktion von $B_{|D}$ erfolgt induktiv: Sei $(Z_t)_{t\in\mathbb{N}}$ eine Familie von unabh. $\nu_{0,1}-$verteilten ZV definiert auf $\Omega = \mathbb{R}^D$, $\mathcal{B}\left(\mathbb{R}^D\right)$, $P(d\omega) = \prod_{i\in D} \nu_{0,1}(d\omega_i)$

$$\begin{aligned} B_0 &:= 0 \\ B_1 &:= Z_1 \end{aligned}$$

auf D_0.

$$B_{\frac{1}{2}} := \frac{1}{2}(B_0 + B_1) + \frac{1}{2} Z_{\frac{1}{2}}$$

auf $D_1 \backslash D_0$. Dann sei also $B_{|D^{n-1}}$ definiert. Auf $t \in D_n \backslash D_{n-1}$:

$$B_t = \frac{1}{2}\left(B_{t-\frac{1}{2^n}} + B_{t+\frac{1}{2^n}}\right) + \frac{1}{2^{\frac{n+1}{2}}} Z_t$$

definiert $B_{|D}$. Beh. $B_{|D}$ erfüllt die Eigenschaften 1. und 2. (Induktion): Betrachte erst $n = 0$ und $n = 1$ mit Zuwächsen auf D_1.

$$\begin{aligned} X_1 &:= B_{\frac{1}{2}} - B_0 = \frac{1}{2}(B_1 - B_0) + \frac{1}{2} Z_{\frac{1}{2}} \\ X_2 &:= B_1 - B_{\frac{1}{2}} = \frac{1}{2}(B_1 - B_0) - \frac{1}{2} Z_{\frac{1}{2}} \end{aligned}$$

Dann sind X_1 und X_2 reell normalverteilt und $\mathbb{E}(X_{1,2}) = 0$ und es folgt:

$$\operatorname{Kov}(X_1, X_2) = \mathbb{E}(X_1 X_2),$$

denn:

$$\begin{aligned} \mathbb{E}(X_1 X_2) &= \mathbb{E}\left(\left[\frac{1}{2}Z_1 + \frac{1}{2}Z_{\frac{1}{2}}\right]\left[\frac{1}{2}Z_1 - \frac{1}{2}Z_{\frac{1}{2}}\right]\right) \\ &\overset{\text{3.bin.F.}}{=} \mathbb{E}\left(\frac{1}{4}Z_1^2\right) - \mathbb{E}\left(\frac{1}{4}Z_{\frac{1}{2}}^2\right) \\ &= \frac{1}{4} - \frac{1}{4} = 0 \end{aligned}$$

Also sind X_1 und X_2 unabh.. Setze für den IS $n-1 \mapsto n$. Seien $s, t \in D_n \backslash D_{n-1}$ mit

$$|s - t| \leq \frac{1}{2^{n-1}}, t \leq s$$

Definiere

$$\begin{aligned}\underline{u} &:= \min\{w\in D_{n+1}\mid w\geq u\}\\ \overline{u} &:= \max\{w\in D_{n+1}\mid w\leq u\}\end{aligned}$$

$$\begin{aligned}\Delta_1 : &= B_{\overline{s}} - B_t\\ \Delta_2 : &= B_s - B_{\overline{s}}\end{aligned}$$

$$\begin{aligned}\Longrightarrow \Delta_1 &= B_{\overline{s}} - \frac{1}{2}\left(B_{\overline{s}} + B_{\overline{t}}\right) - \frac{1}{2^{\frac{n+1}{2}}} Z_t\\ \Delta_2 &= \frac{1}{2}\left(B_{\underline{s}} + B_{\overline{s}}\right) + \frac{1}{2^{\frac{n+1}{2}}} Z_s - B_{\overline{s}}\end{aligned}$$

$\overrightarrow{X} := (\Delta_1, \Delta_2) \in \mathbb{R}^d$ ist Gauß-verteilt in $\mathbb{R}^2$, $\mathbb{E}\left(\overrightarrow{X}\right) = \overrightarrow{0}$.

$$\mathbb{E}\left(\Delta_1\Delta_2\right) = \mathbb{E}\left(\left[\underbrace{\frac{1}{2}\left(B_{\overline{s}} - B_{\overline{t}}\right)}_{=u} - \underbrace{\frac{1}{2^{\frac{n-1}{2}}} Z_t}_{v}\right]\left[\underbrace{\frac{1}{2^{\frac{n+1}{2}}} Z_s}_{w} + \underbrace{\frac{1}{2}\left(B_{\underline{s}} - B_{\overline{s}}\right)}_{y}\right]\right)$$

Dann:

$$\mathbb{E}\left(uw\right) = \mathbb{E}(u)\mathbb{E}(w) = 0$$

nach Konstruktion. Analog gilt $\mathbb{E}(vy) = 0$

$$\mathbb{E}(uy) \overset{IV}{=} \mathbb{E}(u)\mathbb{E}(y) = 0 \Rightarrow \mathbb{E}\left(\Delta_1\Delta_2\right) = 0$$

Zudem sind Z_t und Z_s unabh. Dann:

$$\mathbb{E}(vw) = \mathbb{E}(v)\mathbb{E}(w) = 0 \Rightarrow \mathbb{E}\left(\Delta_1\Delta_2\right) = 0$$

□

$B_{|D}$ hat die gewünschten Eigenschaften. Beweis mittels Induktion.

Beweis. IS: (Δ_1, Δ_2) ist Gauß´sch. Dann gilt:

$$\mathbb{E}\left(\Delta_1\right) = \mathbb{E}\left(\Delta_2\right) = 0$$

und

$$\mathbb{E}\left(\Delta_1\Delta_2\right) = 0$$

da bereits

$$\mathbb{E}\left(D_1 D_2\right) = 0$$

Dann sind die Zuwächse für zwei benachbarte Zeitschritte auf D_n unabh.. Analog für endl. Menge von Zuwächsen sind $\Delta_1, ..., \Delta_n$ unabh. Mit (u, v, w) unabh. ist $(u + v, w)$ unabh und für endl. Zeitschritte der Zuwächse unabh., d.h. falls $0 \leq t_1 \leq ... \leq t_n \leq 1$. Dann:

$$A_i \in D_k \forall i = 1, ..., k$$

und dann:

$$\left(B_{t_n} - B_{t_{n-1}}, ..., B_{t_1} - B_0\right) \text{ ist unabh.}$$

Behaupte dann: $\mathbb{V}\left(B_t - B_s\right) = t - s \forall t, s \in D_n$. Beweis wieder mittels Induktion. □

Wir können $(B_t)_{t\in[0,1]}$ fortsetzen durch

$$\hat{B}_t = \begin{cases} B_t^{(1)} & ,t \in [0,1] \\ B_{t-1}^{(0)} - B_1^{(1)} & ,t \in (1,2] \end{cases}$$

mit $\left(B_t^{(1)}\right)_{t\in[0,1]}$ und $\left(B_t^{(2)}\right)_{t\in[0,1]}$ unabh. Realisierungen von der BB.

Lemma 174. *$X \cong \nu_{m,v_1}, Y \cong \nu_{m_2,v_2}, X, Y$ unabh.. Dann:*

$$X + Y \cong \nu_{m_1+m_2,v_1+v_2}$$

Beweis.

$$\begin{aligned} \mathbb{E}\left(e^{t(X+Y)}\right) &= \mathbb{E}\left(e^{tX}e^{tY}\right) \\ &= \mathbb{E}\left(e^{tX}\right)\mathbb{E}\left(e^{tY}\right) \\ &= e^{tm_1+\frac{t^2}{2}v_1}e^{tm_2+\frac{t^2}{2}v_2} \\ &= e^{t(m_1+m_2)+\frac{t^2}{2}(v_1+v_2)} \end{aligned}$$

□

Zusatz: Falls $X \cong \nu_{0,1}$, dann:

$$P\left(|X| > c\right) \leq \frac{2}{\sqrt{2\pi}}e^{-\frac{c^2}{2}}, c > 1$$

Beweis.

$$\begin{aligned} P\left(|X| > c\right) &= \frac{2}{\sqrt{2\pi}}\int_c^\infty e^{-\frac{t^2}{2}}dt \\ &\leq \frac{2}{\sqrt{2\pi}}\int_c^\infty te^{-\frac{t^2}{2}}dt \\ &= \frac{2}{\sqrt{2\pi}}\left[e^{-\frac{t^2}{2}}\right]_c^\infty \\ &= \frac{2}{\sqrt{2\pi}}e^{-\frac{t^2}{2}} \end{aligned}$$

□

Satz 175. *Die BB ist ein Martingal bzgl. $\mathcal{J}_t = \mathcal{J}_t^{B.}$.*

Beweis. Sei $s \leq t$.

$$\begin{aligned} \mathbb{E}\left(B_t \mid \mathcal{J}_s\right) &= \mathbb{E}\left(B_t \mid \sigma\left(B_u; u \leq s\right)\right) \\ &= \mathbb{E}\left(B_t - B_s + B_s \mid \sigma(.)\right) \\ &= \mathbb{E}\left(B_s \mid \sigma(.)\right) + \mathbb{E}\left(B_t - B_s \mid \sigma(.)\right) \end{aligned}$$

B_s ist $\mathcal{J}_s$-messbar:

$$\begin{aligned} &= B_s + \mathbb{E}\left(B_t - B_s \mid \sigma(.)\right) \\ &\overset{2)}{=} B_s + \mathbb{E}\left(B_t - B_s\right) \\ &= B_s \end{aligned}$$

2) bedeutet, dass der Zuwachs unabh. ist. Die Erwartung ist dann Null, da normalverteilt. □

Bemerkung 176. $X_1, ..., X_k; f : \mathbb{R}^k \to \mathbb{R}^k$ messbar und invertierbar.

$$\overrightarrow{Y} = f(X_1, ..., X_k) = \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix}$$

Dann:

$$\sigma(X_1, ..., X_k) = \sigma(Y_1, ..., Y_k)$$

Hier: $B_{t_1}, ..., B_{t_N}$:

$$\sigma(B_{t_N}, ..., B_{t_1}) = \sigma\left(B_{t_N} - B_{t_{N-1}}, ..., B_{t_1} - B_{t_0}\right)$$

$$f\begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 - X_1 \\ \vdots \\ X_N - X_{N-1} \end{pmatrix}$$

Definition 177. $(X_i)_{i \in I}; X_i : \Omega \to \mathbb{R}$ stoch. Prozess. Sei $J = (j_1, ..., j_M) \subset I$ eine endl. TM.

$$X_J(.) = \begin{pmatrix} X_{j_1}(.) \\ \vdots \\ X_{j_M}(.) \end{pmatrix} : \Omega \to \mathbb{R}^M$$

ist messbar. Dann:

$$P_{X_j} = \mathrm{Vert}(X_J)$$

ist W-Maß auf $\mathbb{R}^M$. $(P_{X_J})_{J \subset I}$ endl."Familie der endl.-dim. Verteilungen des Prozesses $(X_i)_{i \in I}$.

4.4. **Gauss-Prozesse.**

Definition 178. $(X_i)_{i \in I}$ heißt Gauß-Prozess, falls die endl. dim. Verteilungen von X multinomial Gauß´sch sind.

Bemerkung 179. Das W-Maß $(X_{i \in I})$ auf $\mathbb{R}^I = \{\omega \mid \omega : I \to \mathbb{R}\}$ ist bereits durch die endl. dim. Verteilungen von $(X_j)_{j \in J}$ festgelegt.

Satz 180. *(Charakterisierung der BB als Gauß-Prozess) $t \to B_t$ ist eine BB, gdw.:*

(1) $t \to B_t$ ist stetig.
(2) $(B_t)_{t \in \mathbb{R}}$ ist ein Gauß-Prozess mit

$$\mathrm{Kov}(B_s, B_t) = \min(s, t)$$

und

$$\mathbb{E}(B_t) = 0$$

Erinnerung: $Z \in \mathbb{R}^d$ heißt Gaußverteilt, falls

$$Z \cong \nu_{\overrightarrow{m}, A}$$

mit A als Kovarianzmatrix und $\overrightarrow{m} \in \mathbb{R}^d$ als Vektor der Erwartungswerte.

$$(X_{j_1}, ..., X_{j_M}) = Z \in \mathbb{R}^M$$

Dann:

$$\overrightarrow{m} = \begin{pmatrix} \mathbb{E}\left(X_{j_1}\right) \\ \vdots \\ \mathbb{E}\left(X_{j_M}\right) \end{pmatrix};\ A = (A_{\lambda\mu})$$

mit $A_{\lambda\mu} = \text{Kov}\left(X_{j_\lambda}, X_{j_\mu}\right)$. Für die Beschreibung eines Gauß-Prozesses benötigt man:

$$m : I \to \mathbb{R},\ i \mapsto \mathbb{E}\left(X_i\right)$$

die Erwartungsfunktion und

$$K : I \times I \to \mathbb{R}, (i,j) \mapsto \text{Kov}\left(X_i, X_j\right)$$

die Kovarianzfunktion.

Beweis. Die Stetigkeit der Trajektorien wird in beiden Seiten gefordert. Z.z. ist dann: 2) $\iff$ Zuwächse von (B_t) sind normalverteilt und unabh.. Zur Rückrichtung: Sei $\overrightarrow{Z} := (B_{t_N}, ..., B_{t_1})$ normalverteilt, denn $\left(B_{t_N} - B_{t_{N-1}}, ..., B_{t_1} - B_{t_0}\right) =: \overrightarrow{X}$ ist Vektor von unabh. normalvert. ZV. Dann:

$$\overrightarrow{Z} = L^{-1}\overrightarrow{X}$$

ist also wieder multinomial Gauß-vert..

$$m(t) := \mathbb{E}\left(B_t\right) = \mathbb{E}\left(B_t - B_0\right) = 0$$

also ist m eine Nullfunktion.

$$\begin{aligned}
\text{Kov}\left(B_t, B_s\right) &= \mathbb{E}\left(\left[B_t - \mathbb{E}\left(B_t\right)\right]\left[B_s - \mathbb{E}\left(B_s\right)\right]\right) \\
&= \mathbb{E}\left(B_t B_s\right) \\
&= \mathbb{E}\left(\left[B_t - B_s\right] B_s\right) + \mathbb{E}\left(B_s^2\right) \\
&= \mathbb{E}\left(\left[B_t - B_s\right]\left[B_s - B_0\right]\right) + \mathbb{E}\left(\left[B_s - B_0\right]^2\right) \\
&= \mathbb{E}\left(B_t - B_s\right)\mathbb{E}\left(B_s - B_0\right) + s \\
&= s \\
&= \min(t,s) \\
&= K(t,s)
\end{aligned}$$

Die Umkehrung folgt aus der Cholesky-Zerlegung. □

Beispiel 181. $I = \mathbb{R}_{\geq 0}, X_t := B_t, t \mapsto B_t$ BB. Dann $K(i,j) \in \mathbb{R}$ die Kovarianzfunktion und $m(i) \in \mathbb{R}$ die Erwatrtungsfunktion mit

$$K(i,j) = \text{Kov}\left(X_i, X_j\right), m(i) = \mathbb{E}\left(X_i\right)$$

$K : I \times I \to \mathbb{R}$, $m : I \to \mathbb{R}$ entspricht den endl. dim. Verteilungen $\left(X_i\right)_{i \in I}$, die durch $K(.,.)$ und $m(i)$ eind. best. sind.

Bemerkung 182. $K(.,.)$ ist symmetrisch und pos. definit im dem Sinne, dass für alle $J \subset I$ endl. und $\xi \in \mathbb{R}^{(J)}$ gilt:

$$\sum_{k,l \in J} \xi_{kk}(k,l), \xi_l \geq 0$$

denn $Z := \sum_k X_k \xi_k \in \mathbb{R}$. Dann:

$$\begin{aligned}
\mathbb{V}(Z) &\geq \mathbb{E}\left((Z - \mathbb{E}(Z)))^2\right) \\
&= \mathbb{E}\left(Z^2\right) \\
&= \mathbb{E}\left(\left[\sum_{k \in J} X_k \xi_k\right]\left[\sum_{l \in J} X_l \xi_l\right]\right) \\
&= \sum_{k,l \in J} \mathbb{E}\left(\xi_k \xi_l X_k X_l\right) \\
&= \sum_{k,l \in J} \xi_k \xi_l \mathrm{Kov}\left(X_k, X_l\right) \\
&\geq 0
\end{aligned}$$

Satz 183. *Sei I eine Indexmenge und $m : I \to \mathbb{R}$ eine Funktion und $K : I \times I \to \mathbb{R}$ pos. definit im Sinne der letzten Bemerkung. Dann ex. ein Gauß-Prozess $(X_i)_{i \in I}$ mit Werten in $\mathbb{R}$ mit der Erwartungswertfunktion m und Kovarianzfunktion K. (Es ex. $(\Omega, \mathcal{J}, P)$ W-Raum und $(X_i)_{i \in I}$ mit $X_i : \Omega \to \mathbb{R}$ ZV für alle $i \in I$, s.d. (X_i) Gauß-Prozess mit m und K ist.*

Beweis. Erinnerung (kanonischer Pfadraum) I Indexmenge ("Zeitparameter"), $\Omega = \{\omega : I \to \mathbb{R}\} = \mathbb{R}^{\pm}$ (Menge der Abbildungen von I nach $\mathbb{R}$), $X_i : \Omega \to \mathbb{R}$ mittels $X_i(\omega) = (\omega_i)_{i \in I}$, s.d. eine Projektion $\omega_i = \mathrm{pr}_i(\omega)$ ex., $\mathcal{J} := \sigma\left(X_J : \Omega \to \mathbb{R}^{|J|} \mid J \subset I \text{ endl.}\right)$ Wir wollen nun ein geeignetes W-Maß P auf diesem Raum $(\Omega, \mathcal{J})$ ermittlen. Betrachte dazu den Halbring der Zylindermengen

$$\mathcal{Z} = \left\{\mathbb{R} \times ... \times \mathbb{R} \times A_{i_1} \times \mathbb{R} \times ... \times \mathbb{R} \times A_{i_2} \times ... \times A_{i_{|J|}} \times \mathbb{R} \times ...\right\}$$

mit $A_{i_k} \in \mathcal{B}(\mathbb{R})$ für $k \in J = \left\{i_1, ..., i_{|J|}\right\} \subset I$ endl. und dem Inhalt:

$$P_0\left(\mathbb{R} \times ... \times \mathbb{R} \times A_{i_1} \times ... \times A_{i_{|J|}} \times \mathbb{R} \times ...\right) = \nu_{m_J, K_j} \underbrace{\left(A_{i_1} \times ... \times A_{i_{|J|}}\right)}_{\in \mathbb{R}^{|J|}}$$

wobei ν_{m_j, K_j} dem Gauß-Maß im $\mathbb{R}^{|J|}$ entspricht mit dem Erwartungswert $m_j = \left(m\left(i_1\right), ..., m\left(i_{|J|}\right)\right)^T$ und Kovarianzmatrix $K_j = \left(K\left(i_l, i_k\right)\right)_{1 \leq l, k \leq |J|}$ (pos. definit). $P_0(Z)$ ist damit wohldefiniert für alle $Z \in \mathcal{Z}$ unabh. von der genauen Darstellung von Z. Daher ist P_0 ein Prämaß, also ex. eine eind. best. Maßfortsetzung P von $\mathcal{Z}$ auf $\mathcal{J}$ nach Caratheodory. □

Bemerkung 184. $P =$ " $\lim_{|J| \to \infty} \nu_{m_J, K_J}$" ist der inverse Limes, was der Konstruktion eines unendl. dim. Objektes unter Verwendung seiner endl. dim. Projektionen entspricht.

Definition 185. Zwei Prozesse $(X_i)_{i \in I}$ mit $X_i : \Omega \to \mathbb{R}$ und $(Y_i)_{i \in I}$ mit $Y_i : \tilde{\Omega} \to \mathbb{R}$ heißen äquivalent zueinander, falls ihre endl. dim. Projektionen identisch sind. D.h. für alle $J \subset I$ endl. gilt:

$$\mathrm{Vert}\left(X_J\right) = \mathrm{Vert}\left(Y_J\right)$$

als W-Maße auf $\mathbb{R}^{|J|}$.

Damit erhalten wie Äquivalenzklassen von stochastischen Prozessen. Einen Repräsentanten einer solchen Äquivalenzklasse wollen wir Modell nennen. Falls beispielsweise $\Omega = \mathbb{R}^I, X_i := \mathrm{pr}_i(\omega)$ sind, so spricht man vom kanonischen Modell. Umgekehrt: $(Y_i)_{i\in I}, Y_i : \tilde{\Omega} \to \mathbb{R}$ induziert eine Pfadabbildung $\tilde{\omega} \to (Y_i(\tilde{\omega}))_{i\in I} \in \mathbb{R}^I = \Omega$. Dann ist $Y_. : \tilde{\Omega} \to \Omega = \mathbb{R}^I$ eine messbare Abbildung $\left(\tilde{\Omega}, \mathcal{J}\right) \to (\Omega, \mathcal{J})$. Dann wird ein Bildmaß induziert: $\tilde{P} \overset{Y_.}{\to} \hat{P} = \mathrm{Vert}_{\tilde{P}}(Y_.)$. Dann ist $\left(\left(\Omega, \mathcal{J}, \hat{P}\right), X_i : \Omega \to \mathbb{R}, X_i = \mathrm{pr}_i(\omega)\right)$ ein abgeleitetes kanonisches Modell. Zudem ist nach Caratheodory jedes W-Maß auf dem kanonischen Pfadraum durch seine endl. dim. Verteilungen festgelegt.

4.5. **Invarianz der Brownschen Bewegung unter Transformationen.**

Satz 186. *Falls $(B_.)$ eine BB mit $B_0 = 0$ ist, so auch $(B_{t+s} - B_s)_{t\geq 0}, (-B_t)_{t\geq 0}$ und $t \mapsto cB_{\frac{t}{c^2}}$ für $c > 0$, sowie:*

$$X_t := \begin{cases} 0 & t = 0 \\ t - B_{\frac{1}{t}} & t > 0 \end{cases}$$

Beweis. 1) Die endl. dim. Vertielungen von $(B_{t+s} - B_s)_t$ sind Linearkombinationen von endl. dim. Verteilungen von B_t. Also ist auch $(B_{t+s} - B_s)$ ein Gaußscher Prozess und zeitstetig, da (B_t) zeitstetig. Weiterhin gilt:

$$\begin{aligned}
\mathrm{Kov}\left([B_{u+s} - B_s], [B_{r+s} - B_s]\right) &= \mathrm{Kov}(B_{u+s}, B_{r+s}) - \mathrm{Kov}(B_{r+s}, B_s) \\
&\quad -\mathrm{Kov}(B_{u+s}, B_s) + \mathrm{Kov}(B_s, B_s) \\
&= \min(u+s, r+s) - \min(r+s, s) \\
&\quad - \min(u+s, s) + s \\
&= \min(u+r) + s - s - s + s = \min(u, r)
\end{aligned}$$

2) ist klar.
3) Start in 0, gaußsch und zeitstetig und

$$\mathrm{Kov}\left(cb_{\frac{t}{c^2}}, cB_{\frac{s}{c^2}}\right) = c^2 \min\left(\frac{t}{c^2}, \frac{s}{c^2}\right) = \min(t, s)$$

4) Start in 0, $t \mapsto tB_{\frac{1}{t}}$ stetig in $\mathbb{R}_{>0}$, gaußsch für $t \geq 0$ und

$$\begin{aligned}
\mathrm{Kov}\left(u \cdot B_{\frac{1}{u}}, r \cdot B_{\frac{1}{r}}\right) &= ur \min\left(\frac{1}{u}, \frac{1}{r}\right) \\
&= \frac{ur}{\max(u, r)} \\
&= \min(u, r)
\end{aligned}$$

Dann ist

$$t \to \begin{cases} 0 & t = 0 \\ tB_{\frac{1}{t}} & \text{sonst} \end{cases}$$

stetig in $t = 0$. Es ist $(B_t)_{t\in\mathbb{Q}^+}$ wieder gaußsch (mit $\mathbb{Q}^+$ als Parametrisierung. Dann ist $(X_t)_{t\in\mathbb{Q}^+}$ gaußsch. Weiter gilt für alle $Q \subset \mathbb{Q}^+$ endl. ist $(B_t)_{t\in Q}$ gaußsch (und somit auch $(X_t)_{t\in Q}$) jeweils mit $\mathrm{Kov}(s,t) = s \wedge t$, d.h. $\mathrm{Vert}\left((B_t)_{t\in Q}\right) =$

Vert $\left((X_t)_{t\in Q}\right)$ als Maße auf $\mathbb{R}^{|Q|}$. Unter Anwendung des inversen Limes folgt: Vert $\left((X_t)_{t\in\mathbb{Q}^+}\right) = $ Vert $\left((B_t)_{t\in\mathbb{Q}^+}\right)$ als W-Maße auf $\mathbb{R}^{\mathbb{Q}^+} = \{\omega \mid \omega^+ : \mathbb{Q}_+ \to \mathbb{R}\}$, d.h. falls $A \subset \mathbb{R}^{\mathbb{Q}^+}$ messbar, dann gilt:

$$P\left((B_t)_{t\in\mathbb{Q}^+} \in A\right) = P\left((X_t)_{t\in\mathbb{Q}^+} \in A\right)$$

(typische Mengen aus dem Pfadraum sind Zylindermengen: $A = \{\omega \in \mathbb{Q}^+ \mid \omega_{t_1} \in I_1, \omega_{t_2} \in I_2\}$. Wegen der Stetigkeit von $t \to B_t$ folgt:

$$P\left(\lim_{t\to 0} B_t = 0\right) = 1$$

wobei $A = \{\omega \mid \omega : \mathbb{Q}^+ \to \mathbb{R} \mid \lim_{t\to 0} \omega(t) = 0, t \in \mathbb{Q}^+\}$, gdw.

$$P\left(X_{\cdot|\mathbb{Q}^+} \in A\right) = 1$$

s.d.

$$\lim_{t\to 0} X_t(\omega) = 0$$

P-f.s. für $t \in \mathbb{Q}^+$. Dann gilt:

$$\lim_{t\to 0} X_t(\omega) = 0$$

P-f.s.. Damit ist $(X_t)_{t\geq 0}$ eine BB als stetiger Gauß-Prozess mit der Kovarianzfunktion $\Gamma(s,t) = \min(s,t)$. $\square$

4.6. **Die Brownsche Bewegung als Markov-Prozess.** Wir beginnen diesen Abschnitt mit einer längeren Vorbemerkung über die Möglichkeit, die Brown'sche Bewegung in einem beliebigen (bzw. zufällig gewählten) Punkt $x \in \mathbb{R}$ starten zu lassen.

Es sei (B_t) eine standard Brownsche Bewegng und $x \in \mathbb{R}$, dann definieren wir

$$B_t^x = x + B_t$$

als "Brownsche Bewegng mit Start in $x \in \mathbb{R}$". Der Prozess $t \to B_t^x$, $t \geq 0$, ist dann über die folgenden Eigenschaften charakterisiert:

- stetige Trajektorien
- Zuwächse unabh.
- Zuwächse normalverteilt und Varianz entspricht den Zeitschritten
- Start in $x : B_0^x = x$ fast sicher.

Wir betrachten ferner $P_0 :=$ Vert$(B_\cdot)$ (Pfadmaß auf dem kanonischen Pfadraum $\Omega = \mathbb{R}_{\geq 0}$, induziert durch die Standard-BB), welches durch Festlegung auf Zylindermengen $Z \in \mathcal{Z}$ der Form

$$Z = \bigcap_{t_1,\ldots,t_N} P^{-1} Z_j\left(I_j\right) \subset \Omega$$

eindeutig bestimmt ist durch

$$P_0(Z) = P\left(B_{t_1} \in I_N, \ldots, B_{t_N} \in I_N\right)$$

Entsprechend erhalten wir P_x als das Pfadmaß auf $\Omega = \mathbb{R}_{\geq 0}$, das von $(B_t^x)_{t\geq 0}$ induziert und eindeutig bestimmt ist durch best. durch:

$$P_x(Z) = P\left(B_{t_1}^x \in I_1, \ldots, B_{t_N}^x \in I_N\right)$$

Ingesamt erhalten wir somit Familie von Wahrscheinlichkeitsmaßen $\mathbb{R} \times \mathcal{F} \to [0,1]$ mittels $x \times A \to P_x(A)$ wobei $\mathcal{F} = \mathcal{B}(\mathbb{R}_{\geq 0})$ als W-Maße auf dem Pfadraum $(\Omega, \mathcal{F} = \sigma(Z \mid Z \in \mathcal{Z}))$ mit $x \to P_x(A)$ ist stetig (also insb. messbar), $A \in \mathcal{F}$ und $A \to P_x(A)$ ist ein W-Maß auf Ω.

Schließlich können wir nun die Brownsche Bewegung auch in einem zufällig gewählten Punkt $x \in \mathbb{R}$ starten: Sei hierzu ν ein W-Maß auf $\mathbb{R}$ entspechend der W-Verteilung für einen zufällig gewählten Startpunkt. Dann ist die *Mischung*

$$P_\nu(A) = \int_{\mathbb{R}} P_X(A)\nu(dx)$$

wieder ein Maß auf dem kanonischen Pfadraum Ω, welches dem Pfadmaß einer Brown'schen Bewegung mit zufälligem Startwert entsprechend der Startverteilung ν entspricht.

Sei $(X_t)_{t\geq 0} : \Omega \to \mathbb{R}$ der Koordinatenprozess auf Ω, d.h. $X_t(\omega) = \omega(t)$. $(\Omega, \mathcal{F}, P_\nu), (X_t)$. Dann hat (X_t) unter als reellwertiger Prozess, definiert auf dem Maßraum $(\Omega, \mathcal{F})$, die folgenden Eigenschaften

$$\begin{cases} t \to X_t & \text{ist stetig } P_\nu\text{-fast sicher} \\ t \to X_t & \text{hat unabhängige normalverteilte Zuwächse} \\ \operatorname{Vert}(X_0) = \nu & \end{cases}$$

Der Spezialfall einer deterministisch im Punkt $x \in \mathbb{R}$ startenden Brown'schen Bewegung entspricht dann der Wahl von $\nu = \delta_x$.

Der folgende Satz zeigt die Markov'sche Struktur der Brown'schen Bewegung auf. Im Unterschied zu Markov-Ketten sind hier jedoch sowohl die Zeitachse als auch der Zustandsraum kontinuierlich, d.h. $\mathbb{R}_{\geq 0}$ bzw. $\mathbb{R}$.

Satz 187. *(Markoveigenschaft der BB) Sei $(B_t)_{t\geq 0}$ eine BB mit entsprechend ν und $A \in \mathcal{B}(\mathbb{R})$. Dann gilt für $0 \leq t_1 \leq \cdots \leq t_{n+1}$*

$$\mathbb{E}_\nu\left(1_A\left(B_{t_{n+1}}\right) \mid \sigma\left(B_{t_0}, ..., B_{t_n}\right)\right) = P_{B_{t_n}}\left(X_{t_{n+1}-t_n} \in A\right) \textit{ fast sicher.}$$

Bemerkung 188.

(1) Nach dem Faktoriserungslemma für die bedingte Erwartung erwartet man ganz allgemein für die linke Seite in der obigen Gleichung eine Zufallsvariable der Form

$$\mathbb{E}_\nu\left(1_A\left(B_{t_{n+1}}\right) \mid \sigma\left(B_{t_0}, ..., B_{t_n}\right)\right) = F\left(B_{t_0}, ..., B_{t_n}\right)$$

mit einer gewissen messbaren Funktion $F : \mathbb{R}^n \to \mathbb{R}$. Die Identität oben besagt jedoch, dass sich diese Zufallsvariable in der Form

$$\tilde{F}\left(B_{t_n}\right) \text{ mit } \tilde{F} : \mathbb{R} \to \mathbb{R}, \tilde{F}_x = P_x\left(X_{t_{n+1}-t_n} \in A\right)$$

darstellen lässt. Insbesondere gilt also

$$\mathbb{E}_\nu\left(1_A\left(B_{t_{n+1}}\right) \mid \sigma\left(B_{t_0}, ..., B_{t_n}\right)\right) = \mathbb{E}_\nu\left(1_A\left(B_{t_{n+1}}\right) \mid \sigma\left(B_{t_n}\right)\right).$$

Das ist das Analogon zur Gedächtnislosigkeit von Markov-Ketten.

(2) Im Spezialfall für $f : \mathbb{R} \to \mathbb{R}$ messbar erhält man, dass

$$\mathbb{E}_\nu\left(f\left(B_{t_{n+1}}\right) \mid \sigma\left(B_{t_0}, ..., B_{t_n}\right)\right) = \mathbb{E}_{B_{t_n}}\left(f\left(X_{t_{n+1}-t_n}\right)\right).$$

Beim Beweis von Satz 187 benutzten wir das folgende elementare Lemma.

Lemma 189. *Für* $(\Omega, \mathcal{J}, P), X, Y : \Omega \to \mathbb{R}$ *Zufallsvariablen ist*

$$Z := \mathbb{E}_P\left(X \mid \sigma(Y)\right)$$

eindeutig bestimmt durch die Eigenschaft

$$\mathbb{E}_P\left(Xg(Y)\right) = \mathbb{E}\left(Zg(Y)\right)$$

für alle $g : \mathbb{R} \to \mathbb{R}$ *beschränkt messbar.*

Beweis. "$\Rightarrow$"

$$\begin{aligned}\mathbb{E}_P\left(Xg(Y)\right) &= \mathbb{E}_P\left(\mathbb{E}_P\left(Xg(Y) \mid \sigma(Y)\right)\right)\\ &= \mathbb{E}_P\left(\mathbb{E}_P\left(g(Y) \mid \sigma(Y)\right) g(Y)\right)\end{aligned}$$

"$\Leftarrow$" Wähle g in der Menge $\{1_A \mid A \in \mathcal{B}(\mathbb{R})\}$, $g_A(Y) = 1_{Y^{-1}(A)} \in \sigma(Y)$. □

Beweis von von Satz 187. ObdA (via monotone Klassen) genügt es zu zeigen, dass

$$\begin{aligned}&\mathbb{E}_\nu\left(f(B_{t_{n+1}})\, g_0\left(B_{t_0}\right) \cdot \ldots \cdot g_n\left(B_{t_n}\right)\right)\\ &\qquad = \mathbb{E}_\nu\left(\mathbb{E}_{B_{t_n}}\left(f\left(X_{t_{n+1}-t_n}\right) g_0\left(B_{t_0}\right) \cdot \ldots \cdot g_n\left(B_{t_n}\right)\right)\right)\end{aligned}$$

für alle $g_0, \ldots, g_n : \mathbb{R}^n \to \mathbb{R}$ beschränkt messbar. Die linke Seite gleicht

$$\mathbb{E}_{\nu,Z}\left(f\left(z + B_{t_n}\right) g_0\left(B_{t_0}\right) \cdot \ldots \cdot g_n\left(B_{t_n}\right)\right)$$

mit einer unabhängigen Zufallsvariablen $Z \sim \nu_{0,t_{n+1}-t_n}$.

$$\begin{aligned}&= \mathbb{E}_\nu\left(\mathbb{E}_Z\left(f\left(z + B_{t_n}\right)\right) g_0\left(B_0\right) \cdot \ldots \cdot g_n\left(B_{t_n}\right)\right)\\ &= \mathbb{E}_\nu\left(\mathbb{E}_{B_{t_n}}\left(f\left(X_{t_{n+1}-t_n}\right)\right) g_0\left(B_{t_0}\right) \cdot \ldots \cdot g_n\left(B_{t_n}\right)\right)\\ &= \text{rechte Seite.}\end{aligned}$$

□

Bemerkung 190 ((Brownsche) Halbkruppe von Markov-Kernen auf $\mathbb{R}$). Für $s \geq 0$ und $x \in \mathbb{R}$ ist

$$P_s(x, \cdot)$$

ein W-Maß auf $\mathbb{R}$.

$$\begin{aligned}P_s(x, A) &= P_x\left(X_s \in A\right)\\ &= \mathrm{Vert}\left(B_s^x\right)[A]\\ &= \frac{1}{\sqrt{2\pi s}} \int_A e^{-\frac{(s-x)^2}{2s}} dx.\end{aligned}$$

Die Abbildung

$$(x, A) \to P_s(x, A)$$

ist (bei festgehaltenem A) stetig in x, somit also insbesondere messbar und (bei festgehaltenem x) ein W-Maß auf $\mathbb{R}$. $P_s(.,.)$ ist ein Beispiel eines sogenannten Markov-Kerns auf dem Zustandsraum $\mathbb{R}$.

5. Zeitstetige Markov-Prozesse

Wie gesehen ist die Brown'sche Bewegung ist ein Markov-Prozess in kontinuierlicher Zeit und einem kontinuierlichem Zustandsraum. Ausgehend von diesem Beispiel soll nun die Konstruktion von allgemeinen Markov-Prozessen in kontinuierlicher Parametrisierung besprochen werden.

5.1. **Markov-Kerne und Faltungshalbgruppen.**

Definition 191. Es sei $(E, \mathcal{E})$ ein messbarer Raum, dann heißt eine Abbildung

$$K : E \times \mathcal{E} \to [0,1],$$

so dass

$$\begin{cases} x \to K(x,A) & \text{ist messbar} \quad \forall A \in \mathcal{E} \\ A \to K(x,A) & \text{ist ein W-Maß auf}(E, \mathcal{E}) \end{cases}$$

ein *Markov-Kern* auf $(E, \mathcal{E})$.

Bemerkung 192. $K(x,A)$ wird häufig interpretiert als eine Wahrscheinlichkeit, von einem Punkt $x \in E$ in eine Menge $A \in \mathcal{E}$ zu springen. Falls $E \subset \mathbb{N}$ diskret ist, entspricht K einer Übergangsmatrix vermöge $K(x, \{y\}) = K_{xy}$.

Die Hintereinanderausführung von zwei Sprüngen entspricht der folgenden Operation.

Definition 193. Die *Faltung* von Markov-Kernen: $K(.,.), L(.,.)$:

$$(K \cdot L)(x,A) = \int_{\mathbb{R}} K(x,dz)L(z,A)$$

Zum besseren Verständnis führen wir hier noch den Begriff der Faltung von zwei Wahrscheinlichkeitsmaßen ein, der bei Summen von unabhängigen Zufallsvariablen natürlich aufritt.

Definition 194. i) Es sei ν ein Borel'sches W-Maß auf $\mathbb{R}$ und $x \in \mathbb{R}$, dann heißt $\nu_x(A) := \nu(A-x)$ die x-Translation von ν. (Hierbei bezeichnet $A - x := \{a - x, a \in A\}$ die Menge, die durch Verschiebung von A um x nach links entsteht.)
ii) Für zwei W-Maße μ, ν auf $\mathbb{R}$ heißt das W-Maß

$$(\mu * \nu)(A) = \int \nu_x(A)\mu(dx), \quad A \in \mathcal{B}(\mathbb{R})$$

die *Faltung* bzw. das *Faltungsprodukt* der Maße μ und ν.

Die Bedeutung des Faltungsproduktes ergibt such aus dem folgenden Sachverhalt

Lemma 195. *Falls $X \sim \mu$ und $Y \sim \nu$ unabhäangige reele ZVen sind, so gilt für die Summe*

$$Z := X + Y \sim (\mu * \nu)$$

Bemerkung 196. Die Faltung und das obige Resultat lassen sich unmittelbar auf den vektorwertigen Fall, d.h. wenn etwa μ und ν W-Maße auf $\mathbb{R}^d$ sind, erweitern.

Beweis von Lemma 195. Es sei $f : \mathbb{R} \to \mathbb{R}$ beschränkt messbar, dann

$$\begin{aligned} \mathbb{E}\left(f(Z)\right) &= \mathbb{E}_X \mathbb{E}_Y \left(f(X+Y)\right) \\ &= \mathbb{E}_X \left(\mathbb{E}_Y (X+Y)\right) \\ &= \int_{\mathbb{R}} \mathbb{E}_Y \left(f(X+Y)\right)\mu(dx) \\ &= \iint_{\mathbb{R}} f(Z)\mathrm{Vert}_{(X+Y)}(dz)\mu(dx) \end{aligned}$$

mit

$$\begin{aligned}\mathrm{Vert}_{(X+Y)}(A) &= P(x+Y\in A)\\ &= P(Y\in A-x)\\ &= \nu(A-x)\\ &= \nu_x(A).\end{aligned}$$

Folglich

$$\begin{aligned}\mathbb{E}\left(f(Z)\right) &= \iint_{\mathbb{R}} f(z)\nu_x(dz)\mu_x(dx)\\ &= \int_{\mathbb{R}} f(z)\left(\int_{\mathbb{R}}\nu_x(dx)\mu(dx)\right)\\ &= \int_{\mathbb{R}} f(z)(\mu*\nu)(dz)\end{aligned}$$

□

Korollar 197. $(\mu*\nu)=(\nu*\mu)=\mathit{Vert}(X+Y)=\mathit{Vert}(Y+X)$

Die Faltung zweier Maße ist eine Verallgemeinerung der Faltung von Funktionen aus der Analysis, an den wir hier erinnern wollen.

Definition 198. Seien $g,h:\mathbb{R}\to\mathbb{R}$ gegeben. Dann:

$$(g*h)(x)=\int_{\mathbb{R}} g(x-y)h(y)dy$$

Falls $\mu(dx)=g(x)dx$ und $\nu(dx)=h(x)dx$, dann:

$$(\mu*\nu)(dx)=(g*h)(x)dx$$

Man macht sich leicht klar, dass für die Faltung zweier absolut stetiger W-Maße $\mu(dx)=m(x)dx$ und $\nu(dx)=n(x)dx$ gilt, dass $(\mu*\nu)(dx)=(m*n)(x)dx$.

Definition 199. Eine Familie $(\nu_s)_{s\geq 0}$ von W-Maßen auf $\mathbb{R}$ heißt Faltungshalbgruppe, falls $(\nu_s*\nu_t)=\nu_{s+t}$ für alle $s,t\geq 0$ gilt.

Definition 200. $(f_s)_{s\geq 0}$ Familie von Funktionen auf $\mathbb{R}$ heißt Faltungshalbgruppe, gdw. $(f_s*f_t)=f_{s+t}$.
Beispiel 201.

(1) Gaußsche Faltungshalbgruppe:

$$f_s(x)=\begin{cases}\frac{1}{\sqrt{2\pi s}}e^{-\frac{x^2}{2s}} & , s>0\\ \infty 1_{\{0\}}(x) & , s\leq 0\end{cases}$$

denn: $f_s(x)dx=\nu_{0,s}(dx)\cong X$, $f_t(x)dx=\nu_{0,t}(dx)\cong Y$. Dann:

$$Z=X+Y\cong(f_t*f_s)(x)dx\cong f_{t+s}(x)dx$$

(Hierbei verwenden wir unsere Kenntnisse über Summen von unabhängigen normalverteilten Zufallsvariablen (siehe WT I).)

(2) Cauchyverteilungen:

$$\gamma_\alpha(dx)=\frac{\alpha}{\pi}\frac{1}{\alpha^2+x^2}dx,\quad \alpha>0$$

WMaß auf $\mathbb{R}$. Dann ist $(\gamma_\alpha)_{\alpha>0}$ eine Faltungsgruppe von Wmaßen auf $\mathbb{R}$ (Beweis: siehe Übungen.)

Die Faltung zweier Markov-Kerne $(K * L)(x, A)$ gibt die Wahrscheinlichkeit an, in einem zweifachen Sprung von x in die Menge A zu springen, wobei der erste Sprung zufällig gemäß K und der zweite Sprung zufällig gemäß L erfolgt. Analog zur Sprechweise für Maße kann man nun auch den folgenden Begriff einführen.

Definition 202. Eine Familie $(K_s(.,.))_{s\geq 0}$ von Markovkernen heißt *Halbgruppe von Markovkernen*, falls

$$(K_s * K_t) = K_{s+t}$$

Dieser Begriff verallgemeinert den von Markovketten bekannten Zusammenhang, dass $P^{(m)} \cdot P^{(n)} = P^{(n+m)}$.

Jede Faltungshalbgruppe von W-Maßen auf $\mathbb{R}$ erzeugt eine Halbgruppe von Markovkernen wie folgt.

Satz 203. *Falls $(\nu_s)_{s>0}$ Faltungshalbgruppe von WMaßen auf $\mathbb{R}$ ist, dann definiert*

$$K_s(x, A) \;=\; \nu_s(A - x)$$

eine Faltungshalbgruppe von Markovkernen.

Beweis.

$$\begin{aligned}
(K_s * K_t)(x, A) &= \int_{z\in\mathbb{R}} K_s(x, dx) K_t(z, A) \\
&= \int_{\mathbb{R}} \nu_S(dz - x)\nu_t(A - z) \\
&= \int_{\mathbb{R}} \nu_s(dz)\nu_t(A - (z + x)) \\
&= \int_{\mathbb{R}} \nu_s(dz)\nu_t(\underbrace{(A - x)}_{\tilde{A}} - z) \\
&= (\nu_s * \nu_t)\left(\tilde{A}\right) \\
&= \nu_{s+t}\left(\tilde{A}\right) \\
&= \nu_{s+t}(A - x) \\
&= K_{s,t}(x, A)
\end{aligned}$$

□

5.1.1. *Konstruktion von zeitstetigen Markov-Prozessen durch Inversen Limes.* In diesem Abschnitt wird nun eine Faltungs-Halbgruppe $(K_s(.,.))_{s>0}$ von Markov-Kernen auf $\mathbb{R}$ für die Konstruktion eines Wahrscheinlichkeitsmaßes auf dem kanonischen Pfadraum $\Omega := \mathbb{R}^I$ mit $I = \mathbb{R}_{\geq 0}$ genutzt. Die Grundidee ist dieselbe wie schon im Fall von Markovketten, bei der man zunächst einen Inhalt auf der Menge der Zylindermengen definiert. Man prüft anschließend die Wohldefiniertheit dieses Inhaltsbegriffs, sowie in einem letzten Schritt die Prämaßeigenschaft. Die Existanz und Eindeutigkeit eines hierdurch festgelegten

Wahrscheinlichkeitsmaßes folgt dann durch Anwendung des Satzes von Caratheodory. Es folgen nun die Schritte im Einzelnen.

1) Definition eines Inhalts auf den Zylindermengen. Es sei ein Punkt $x \in R$ (als 'Startpunkt') fest gewählt.

Sei nun $Z \in \mathcal{Z} \subset 2^{\Omega}$ eine Zylindermenge im kanonischen Pfadraum, d.h. eine Menge $Z \subseteq \mathbb{R}^{\mathbb{R}>0}$ der Form

$$Z = \{\omega \in \mathbb{R}^{\mathbb{R}_{\geq 0}} \,|\, \omega_{t_i} \in A_i, i = 1, \ldots, N\}, \tag{4}$$

wobei $0 \leq t_1 \leq \cdots \leq t_N$ und $A_i \in \mathcal{B}(\mathbb{R})$, $i = 1, \ldots, N$ vorgegebene Beobachtungszeitpunkte bzw. Beobachtungsmengen sind.

Eine Trakektorie $\omega.$ gehört damit zur Zylindermenge Z genau dann, wenn sie die Folge von 'Fenstermengen' $A_i \subset \mathbb{R}$, $i \in 1, \ldots, N$ in den aufeinanderfolgenden Zeitpunkten $0 \leq t_1 \leq t_2 \cdots \leq t_N$ durchläuft. Entsprechend unserer Vorstellung, dass die Familie der Markovkerne $K_s(.,.)$ die Wahrscheinlichkeiten für Übergänge der Form $P(X_{t+\Delta t} \in A|X_t = x) = P_{\Delta t}(x, A)$ beschreibt, ergibt sich der folgende Ansatz für die Wahrscheinlichkeit, eine zur Zylindermenge Z zugehörige Trajekorie zu beobachten wie folgt

$$\begin{aligned} P_x(Z) = \int_{x_1 \in A_1} K_{t_1}(x, dx_1) \int_{x_2 \in A_2} K_{t_2 - t_1}(x_1, dx_2) \ldots \\ \ldots \int_{x_{N-1} \in A_{N-1}} K_{t_{N-1} - t_{N-2}}(x_{N-2}, dx_{N-1}) K_{t_N - t_{N-1}}(x_{N-1}, A_N). \end{aligned} \tag{5}$$

Hierbei verwenden wir (im Sonderfall, dass $t_i = t_{i+1}$) die Konvention, dass

$$K_0(x, A) := \delta_x(A) = 1_A(x).$$

K_0 ist der Markov-Kern, der das deterministische Verharren im jeweiligen Ausgangspunkt eines (somit ausbleibenenden) Sprungs beschreibt. Man macht sich schnell klar, dass hierdurch die Faltungshalbgruppe als $(K_s)_{s \geq 0}$ fortgesetzt wird, d.h. insbesondere gilt $K_0 * K_t = K_t$ für alle $t \geq 0$.

2) Wohldefiniertheit des Inhalts P_x auf den Zylindermengen.
Man beachte, dass die Darstellung von Z in (4) nicht eindeutig ist. In der Tat, für jede Verfeinerung $\mathcal{J} = \{t'_1, \ldots, t'_M\} \supset \mathcal{I} := \{t_1, \ldots, t_N\}$ von (Beobachtungs-)Zeitpunkten und (Fenster-)Mengen $A'_1, \ldots A'_M$ gilt ebenso

$$Z = \{\omega \in \mathbb{R}^{\mathbb{R}_{\geq 0}} \,|\, \omega_{t'_j} \in A'_j, j = 1, \ldots, M\},$$

sofern gewährleistet ist, dass $A'_j = A_i$, falls $t'_j = t_i$ bzw. $A'_j = \mathbb{R}$ für alle j mit $t'_j \in \mathcal{J} \setminus \mathcal{I}$.
Zum Nachweis, dass $P_x(Z)$ durch die Vorschrift (6) dessen ungeachet wohldefiniert ist, beschränken wir uns hier der Einfachheit halber auf den Fall, dass $M = N + 1$ mit $\mathcal{J} = \{t_1 \leq \ldots t_l < \tau < t_{l+1} \leq t_{l+2} \cdots \leq t_N\}$ und der Folge von Fenstermengen $A_1, A_2, \ldots, A_l, \mathbb{R}, A_{l_1}, \ldots, A_N$.

Die Anwendung der Vorschrift (6) liefert in diesem Fall den Ausdruck

(6)

$$P_x(Z) = \int_{x_1\in A_1} K_{t_1}(x,dx_1) \int_{x_2\in A_2} K_{t_2-t_1}(x_1,dx_2) \cdots \int_{x_{l-1}\in A_{l-1}} K_{t_{l-1}-t_{l-2}}(x_{l-2},dx_{l-1})$$
$$\int_{x_l\in A_l} K_{t_l-t_{l-1}}(x_{l-1},dx_l) \int_{z\in\mathbb{R}} K_{\tau-t_l}(x_l,dz) \int_{x_{l+1}\in A_{l+1}} K_{t_{l+1}-\tau}(z,dx_{l+1})$$
$$\int_{x_{l+2}\in A_{l+2}} K_{t_{l+2}-t_{l+1}}(x_{l+1},dx_{l+2}) \cdots \int_{x_{N-1}\in A_{N-1}} K_{t_{N-1}-t_{N-2}}(x_{N-2},dx_{N-1}) K_{t_N-t_{N-1}}(x_{N-1},A_N$$

Hierbei ist nach Definition der Faltung von Markov-Kernen

$$\begin{aligned}\int_{z\in\mathbb{R}} K_{\tau-t_l}(x_l,dz) K_{t_{l+1}-\tau}(z,dx_{l+1}) &= \left(K_{\tau-t_l} * K_{t_{l+1}-\tau}\right)(x_l,dx_{l+1}) \\ &= K_{t_{l+1}-t_{l+1}}(x_l,dx_{l+1}),\end{aligned}$$

wobei wir im letzten Schritt die Halbgruppeneigenschaft der Familie $(K_s)_{s\geq 0}$ ausgenutzt haben. Durch die Vorschrift (6) ist $P_x(Z)$ also wohldefiniert.

3) Prämaßeigenschaft und Anwendung des Fortsetzungssatzes von Caratheodory. Für die Anwendung des Satzes von Caratheodory zur Fortsetzung des Inhaltes P_x auf $\mathcal{Z}$ zu einem Maß auf der von $\mathcal{Z}$ erzeugten Sigma-Algebra $\mathcal{F} = \sigma(\mathcal{Z})$ führen wir jetzt noch eine Sprechweise ein, die den von uns eingeschlagenen Weg zur Konstruktion eines Wahrscheinlichkeitsmaßes auf einem unendlich-dimensionalen Produktraum $\mathbb{R}^I$ mit einer beliebigen Indexmenge $\mathcal{I}$ systematisiert. – Die uns bereits aus der W-Theorie I bekannten Konstruktionen von Modellen für unendliche Folgen von unabhängigen Zufallsvariablen fallen passen dann ebenfalls in diese Systematik genauso wie die unser Existenzresultat für diskrete Markovketten zu gegebener Startverteilung λ und Übergangsmatrix P.

Sei nun $\mathcal{I}$ eine solche feste beliebige Indexmenge und $\Omega := \mathbb{R}^{\mathcal{I}}$ der zugehörige Pfadraum sowie $\mathcal{F} := \sigma(\mathcal{Z}) \subset 2^{\Omega}$ die von den Zylindermengen $\mathcal{Z}$ erzeugte Sigma-Algebra, wobei

$$\Omega \supset Z \in \mathcal{Z} :\Leftrightarrow \exists i_1,\ldots,i_N \in \mathcal{I}, A_i \in \mathcal{B}(\mathbb{R}), i=1,\ldots,N : Z = \cap_{k=1}^N \pi_{i_k}^{-1}(A_k).$$

Hierbei bezeichnet $\pi_i : \mathbb{R}^{\mathcal{I}} \to \mathbb{R}$, $\omega = (\omega_j)_{j\in\mathcal{I}} \to \omega_i$ die Auswertungs- bzw. Koordinatenabbildung zur Koordinate $i \in \mathcal{I}$. Allgemeiner haben wir für jedes Paar von Teilmengen $\mathcal{K} \subset \mathcal{J} \subset \mathcal{I}$ die Projektionsabbildungen

$$\pi_{\mathcal{K}}^{\mathcal{J}} : \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{K}}, (\omega_i)_{i\in\mathcal{J}} \to (\omega_i)_{i\in\mathcal{K}},$$

wobei wir den oberen Index $\mathcal{J}$ nicht schreiben wollen, falls $\mathcal{J} = \mathcal{I}$, so dass also $\pi_i = \pi_{\{i\}}$.

Eine Menge $Z \subset \mathbb{R}^I$ ist somit eine Zylindermenge genau dann, wenn es eine endliche Teilmenge $I \subset \mathcal{I}$ und Borelmengen $A_i \subset \mathbb{R}$, $i \in I$ gibt, so dass $Z = \pi_I^{-1}(A_1 \times \cdots A_N)$. Entsprechend läst sich die Definition (6) in der Form schreiben

$$P_x(Z) = P_x^I(A_1 \times \cdots A_N),$$

wenn P_x^I das Wahrscheinlichkeitsmaß das durch

(7)

$$P_x^I(A_1 \times \cdots A_N) = \int_{x_1 \in A_1} K_{t_1}(x, dx_1) \int_{x_2 \in A_2} K_{t_2 - t_1}(x_1, dx_2) \dots$$
$$\dots \int_{x_{N-1} \in A_{N-1}} K_{t_{N-1} - t_{N-2}}(x_{N-2}, dx_{N-1}) K_{t_N - t_{N-1}}(x_{N-1}, A_N).$$

Die Familie der Maße $(P_x^I)_{I \subset \mathbb{R}_{\geq 0}}$ ist nun *konsistent* in dem folgenden Sinne, sie mit der Familie zugehörigen Projektionen verträglich ist

Definition 204. Es sei $\mathcal{I}$ eine Indexmenge und zu jeder endllichen Teilmenge $I \subset \mathcal{I}$ sei ein W-Maß μ_I auf $\mathbb{R}^I$ gegeben. Dann heißt die Familie $(\mu_I)_{I \subset \mathcal{I}, I endlich}$ *konsistent*, falls für alle $J \subset I$ gilt, dass $(\pi_J^I)_* \mu_I = \mu_J$.

Beispiel 205.

i) Schaut man sich die Diskussion zur Wohldefiniertheit vom Zylindermaß P_x auf $\mathcal{Z}$ noch einmal an, sieht man, dass dort in der Tat $P_x^I(A_1 \times \cdots \times A_N) = P_x^J(A_1 \times \cdots \times A_l \times \mathbb{R} \times A_{l+1} \times \cdots A_N)$ gezeigt wurde, wobei $I = \{t_1, \dots, t_N\}$ und $J = I \cup \{\tau\}$ mit $t_l \leq \tau \leq t_{l+1}$. Dies ist aber gleichbedeutend mit der Aussage, dass $(\pi_I^J)_* P_x^J = P_x^I$, was man leicht auf den allgemeinen Fall $J \subset I$ verallgemeinert. Die Familie $(P_x^I)_{I \subset \mathbb{R}_{\geq 0}}$ ist somit konsistent.

ii) Ein anderes Beispiel einer Konsistenten Familie von Wahrsheinlichkeitsmaßen für $\mathcal{I}$ beliebig entsteht durch $\mu_I(dx_{i_1}, \dots, dx_{i_N}) := \otimes_{k=1}^N \mu(dx_{i_k})$ (N-faches Produktmaß, $I = \{i_1, \dots, i_N\}$), für die spätere Konstruktion einer durch $i \in \mathcal{I}$ parametrisierten i.i.d. Familie von Zufallsvariablen mit Verteilung $X_i \sim \mu$.

Der folgende Satz zeigt nun, dass eine Familie von W-Maßen $(\mu_I)_{I \subset \mathcal{I}}$ im Sinne der Definition 204 genau dann konsistent ist, wenn sie aus den endlichdimensionalen Verteilungen eines eindeutig bestimmten Wahrscheinlichkeitsmaßes $\mu_{\mathcal{I}}$ auf $(\mathbb{R}^{\mathcal{I}}, \mathcal{F} = \sigma(\mathcal{Z}))$ besteht.

Satz 206 (Kolmogorov'scher Konsistenzsatz)**.** *Sei $\mathcal{I}$ eine Indexmenge und $(\mu_I)_{I \subset I, I endlich}$ eine konsistente Familie von Wahrscheinlichkeitsmaßen im Sinne der Definition 204. Dann existiert eine eindeutige Fortsetzung $\mu_{\mathcal{I}}$ als Wahrscheinlichkeitsmaß auf $\mathbb{R}^{\mathcal{I}}$ von dem durch*

$$\mu(Z) := \mu_I\left(A_1 \times \dots \times A_N\right), \quad Z = \pi_I^{-1}(A_1 \times \cdots A_N) \in \mathcal{Z}$$

festgelegten Inhalt auf den Zylindemengen in $R^{\mathcal{I}}$.

Beweis. Die Konsistenz der Familie (μ_I) liefert zunächst, dass der Inhalt $\mu(Z)$ für $Z \in \mathcal{Z}$ durch die obige Vorschrift wohldefiniert ist. Die Existenz und Eindeutigkeit einer Fortsetzung von μ als W-Maß auf der von $\mathcal{Z}$ erzeugten σ-Algebra folgt aus dem Fortsetzungssatz von Caratheodory. Hierzu muss man jedoch überprüfen, dass μ die Prämaß-Eigenschaft hat. Hierzu genügt es, die "Stetigkeit von oben" für μ zu zeigen. Diesen Beweis hatten wir im Falle, dass $\mathcal{I} = \mathbb{N}$ bereits in der Wahrscheinlichkeitstheorie I geführt, vergleiche dort den Beweis von Satz 2.11. Dieser Beweis für beliebige Indexmengen $\mathcal{I}$ ist identisch. □

Bemerkung 207. Aussage und Beweis vom Kolmogorov'schen Konsistenzsatz bleiben gültig, wenn man $\mathbb{R}$ durch einen metrischen Raum (E, d) ersetzt, welcher vollständig und separabel ist, d.h. eine abzählbare dichte Teilmenge enthält (d.h. wenn (E,d) ein sog. *'Polnischer Raum'* ist).

Offensichtlich ist gilt schließlich für die von den Zylindermengen erzuegte σ-Algebra $\mathcal{F} = \sigma(\mathcal{Z}) = \sigma(\pi_s, s \geq 0)$. Wir halten als Zwischenergebnis somit das folgende fest.

Satz 208. *Es sei $(K_s(.,.))_{s\geq 0}$ eine Halbkruppe von Markov-Kernen auf $\mathbb{R}$, dann existiert zu jedem $x \in \mathbb{R}$ ein eindeutig bestimmtes W-Maß auf dem kanonischen Pfadraum $(\Omega = \mathbb{R}^{\mathbb{R}_{\geq 0}}, \mathcal{F} = \sigma(\pi_t, s \in \mathbb{R}))$, so dass die endlichdimensionalen Verteilungen von P_x durch $(P_x^I)_{I\subset\mathbb{R}_{\geq 0}}$ gemäß (7) gegeben sind.*

5.2. **Universeller Markov Prozess.** Führen wir die obige Konstruktion für jedes $x \in \mathbb{R}$ als Startpunkt fest, erhalten wir mit $(P_x)_{x\in\mathbb{R}}$ eine durch die Menge aller Startpunkte parametrisierte Familie von Wahrscheinlichkeitsmaßen auf dem Pfadram Ω. Auf $(\Omega, \mathcal{F})$ ist für $s \geq 0$ die Koordinatenabbildung $X_s : \Omega \to \mathbb{R}$, $\omega = (\omega_s)_{s\geq 0} \to X_s(\omega) := \pi_s(\omega) = \omega_s$ (trivialerweise) messbar. Insgesamt erhalten wir somit ein Quadrupel

$$(\Omega, \mathcal{F}, (P_x)_{x\in\mathbb{R}}, (X_s)_{s\geq 0}),$$

bestehend aus einem messbaren Raum $(\Omega, \mathcal{F})$, einem hierauf definierten stochastischen Prozess $(X_s)_{s\geq 0}$ mit Werten in $\mathbb{R}$ (als sog. *'Zustandsraum'*) und einer durch die Punkte $x \in \mathbb{R}$ des Zustandsraums parametrisierte Famillie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{F})$. Allgemeiner können wir den Zutsandsraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ auch durch einen beliebigen messbaren Räumen $(E, \mathcal{E})$ als Zustandsraum ersetzen (vergl. Bemerkung 207).

Definition 209. Es sei $(E, \mathcal{E})$ ein messbarer Raum. Ein Quadrupel $(\Omega, \mathcal{F}, (P_x)_{x\in E}, (X_s)_{s\geq 0})$, bestehend aus einem messbaren Raum $(\Omega, \mathcal{F})$, einem hierauf definierten E-wertigen stochastischen Prozess $(X_s)_{s\geq 0}$, $X_s : \Omega \to E$ und einer durch die Punkte von E parametrisierten Familie von Wahrscheinlichkeitsmaßen $(P_x)_{x\in E}$ auf $(\Omega, \mathcal{F})$ heißt *universeller Markov-Prozess*, falls

1) $x \to P_x(A)$ ist messbar für alle $A \in \mathcal{F}$

2) $X_0(\omega) = x$ für P_x-f. alle $\omega \in \Omega$

3) $P^x\left(X_{s+t} \in B \mid \mathcal{F}_s^X\right) = P_{X_s}(X_t \in B) \quad P_x$-fast sicher, $\forall s, t > 0, B \subset \mathbb{R}$ messbar

Bemerkung 210. Die Eigenshchaft 3. ist dabei zu lesen als

$$\mathbb{E}_{P_x}\left(1_B\left(X_{s+t}\right) \mid \sigma\left(X_u \mid u \leq s\right)\right) = \mathbb{E}_{P_{X_s}}\left(1_B\left(X_t\right)\right) P_x - \text{fast sicher.} \qquad (*)$$

Dies ist das Analog zur Markov-Eigenschaft im diskreten Fall

$$P_X\left(X_{k+l} \in A \mid X_1, ..., X_k\right) = P_{X_k}\left(X_l \in A\right)$$

Satz 211. *Es sei $(K_s)_{s\geq 0}$ eine Halbgruppe von Markov-Kernen auf $\mathbb{R}$ und $(P_x)_{x\in\mathbb{R}}$ die zugehörige Familie von Wahrscheinlichkeitsmaßen auf dem kanonischen Pfadraum $(\Omega, \mathcal{F})$, sowie $X_s : \Omega \to \mathbb{R}$, $X_s(\omega) := \omega_s$ der kanonische Koordinatenprozess auf Ω. Dann bildet $(\Omega, \mathcal{F}, (P_x)_{x\in\mathbb{R}}, (X_s)_{s\geq 0})$ einen universellen Markov-Prozess.*

Bemerkung 212. Aussage und Beweis gelten entsprechend auch im Fall eines Polnischen Zustandsraums $(E, \mathcal{E})$ anstelle von $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In diesem Fall sind die Markov-Kerne K_s definiert als Abbildungen $K_S : E \times \mathcal{E} \to [0,1]$.

Beweise von Satz 211. Die Eigenschaft 1 folgt aus der aus der Messbarkeit der Vorschrift

$$x \longrightarrow K_s(x,A) \qquad \forall s, A \text{ fest.}$$

Die Aussage 2 ist eine Konsequenz von $K_0(x,.) = \delta_x(.)$, d.h.

$$K_0(x,A) = \delta_x(A) = \begin{cases} 1 & , x \in A \\ 0 & , x \notin A, \end{cases}$$

so dass

$$P_X\left(X_0 \in A\right) = K_0(x,A) = \begin{cases} 1 & , x \in A \\ 0 & , x \notin A. \end{cases}$$

Also

$$\mathrm{Vert}_{P_X}\left(X_0\right) = \delta_x,$$

d.h. unter dem Pfadmaß starten alle Trajektorien x fast sicher.

Zum Beweis der Eigenschaft 3 reicht es (mit masstheoretischer Induktion, d.h. etwa einem montone Klassen-Argument) zu zeigen, dass für $0 \leq s_1 \leq ... \leq s_n = s$ und $f, (g_i)_{i=1,\dots,n}$ beschränkt messbar

$$\mathbb{E}_{P_X}\left(f\left(X_{s+t}\right) g_1\left(X_{s_1}\right) ... g_n\left(X_{s_n}\right)\right) = \mathbb{E}_{P_X}\left(\mathbb{E}_{P_{X_s}}\left(f\left(X_t\right)\right) g_1\left(X_{s_1}\right) ... g_n\left(X_{s_n}\right)\right)$$

Dann gilt: $\mathbb{E}_{P_X}\left(f\left(X_{s+t}\right) \mid \mathcal{J}_s^x\right) = \mathbb{E}_{P_{X_s}}\left(g\left(X_t\right)\right)$. ObdA seien f und (g_i) chrakteristische Funktionen. Wir zeigen die gewünschte Identität im Spezialfall $n = 2$, der allgemeine Fall ergibt sich dann duch Induktion nach n.
Entsrpechend (6) schreibt sich die linke Seite oben in der Form

$$P_X\left(X_{s+t} \in A; X_s \in B; X_u \in C\right) = \int_{X_1 \in C} K_u\left(x, dX_1\right) \int_{X_2 \in B} K_{s-u}\left(x_1, dX_2\right) \overbrace{K_t\left(x_2, A\right)}^{=:h(x_2)}$$

mit $0 \leq u \leq s$ und $h(z) := \mathbb{E}_{P_Z}\left(1_A\left(X_t\right)\right)$. Somit haben wir die behauptete Identität bewiesen, falls

$$(8) \qquad \mathbb{E}_{P_X}\left(h\left(X_s\right) 1_B\left(X_s\right) 1_C\left(X_u\right)\right) = \int_C K_u\left(x, dx_1\right) \int_B K\left(x_1, dx_2\right) h\left(X_2\right)$$

Letztere Gleichung ist aber erneut eine Konsequenz aus der Definiton von P_X und aus dem abgeleiteten Integral bzgl. P_X, denn

$$\begin{aligned} &\mathbb{E}_{P_X}\left(h_1\left(X_{t_1}\right), ..., h_n\left(X_{t_n}\right)\right) \\ &= \int_{x_1 \in \mathbb{R}} K_{t_1}\left(x, dx_1\right) h\left(x_1\right) \int_{x_2 \in \mathbb{R}} K_{t_2 - t_1}\left(x_1, dx_2\right) h_2\left(x_2\right) \\ &\quad ... \int_{x_n \in \mathbb{R}} K_{t_n - t_{n-1}}\left(x_{n-1}, dx_n\right) h_n\left(X_n\right) \end{aligned}$$

Somit ist (8) ein Spezialfall mit $t_1 = u, t_2 = s$ und $h_1 = 1_C, h_2 = 1_B \cdot h$ □

Bemerkung 213. Man beachte, dass im Falle eines universellen Markov-Prozesses der stochastische Prozess (z.B. als Koordinatenprozess) für alle Startpunkte stets derselbe ist. Der Startpunkt bestimmt lediglich das Wahrscheinlichkeitsmaß, unter welchem man den Prozess betrachtet, d.h. also nach welchem Zufallsgesetz die (zu beobachtenden) Trajektorien auswählt werden.

Bemerkung 214 (Markov-Eigenschaft für allgemeine Pfadfunktionale). Mit denselben Argumenten wie oben folgt für einen universellen Markov-Prozess, dass auch

$$E_x(1_{A_1}(X_{t_1+u})\cdots 1_{A_M}(X_{t_M+u})|\mathcal{F}_u^X) = E_{X_u}(1_{A_1}(X_{t_1})\cdots 1_{A_M}(X_{t_M}) \text{ fast sicher,}$$

wobei $0 \leq t_1 \cdots \leq t_M$ und $u \geq 0$. Durch masstheoretische Induktion erhalten wir hiermit schließlich für messbare Funktionale $F : \mathbb{R}^{\mathbb{R}_{\geq 0}} \to \mathbb{R}$ auf dem kanonischen Pfadraum, dass

$$E(F(X_{u+\cdot}|\mathcal{F}_u^X) = E_{X_u}(f(X_\cdot))] \text{ fast sicher,}$$

wobei $X_{u+\cdot} : \Omega \to \mathbb{R}^{\mathbb{R}_{\geq 0}}$, $X_{u+\cdot} = (X_{u+s})_{s\geq 0}$ die um den Zeitschritt u verschobene Version der Trajektorienabbildung $X_\cdot : \Omega \to \mathbb{R}^{\mathbb{R}_{\geq 0}}$, $X_\cdot = (X_s)_{s\geq 0}$ bezeichnet.

5.3. **Starke Markov-Eigenschaft und Spiegelungsprinzip der Brownschen Bewegung.** Analog zur zeitdiskreten Situation kann man nun auch Stoppzeiten etc. einführen und die Frage nach der Markov-Eigenschaft für solche Zufallszeiten setellen.

Definition 215. Es sei $(\Omega, \mathcal{F})$ ein messbarer Raum und $\mathcal{F}_\cdot = (\mathcal{F}_s)_{s\geq 0}$ eine Filtration mit $\mathcal{F}_s \subset \mathcal{F}$ für alle $s > 0$. Dann heißt $\tau : \Omega \to \mathbb{R} \cup \{\infty\}$ eine $\mathcal{F}_\cdot$-Stoppzeit, falls $\{\tau \leq t\} \in \mathcal{F}_t$ für alle $t \geq 0$. In diesem Fall heißt das Mengensystem $\mathcal{F}_\tau := \{A \in \mathcal{F}\,|\,A \cap \{\tau \leq t\} \in \mathcal{F}_t\, \forall t \geq 0\}$ die σ-Algebra der τ-Vergangenheit.

Bemerkung 216. Analog zur zeitdiskreten Situation zeigt man, dass $\mathcal{F}_\tau$ dann in der Tat eine σ-Algebra ist.

Im folgenden wollen von der Situation einer beliebigen (standard) Brown'schen Bewegung ausgehen, definiert als Prozess $(B_t)_{t\geq 0}$ auf einem gewissen Wahrscheinlichkeitsraum (Ω, P). Man beachte, dass wir durch verschiebung des Startpunkte $B_t^x := x + B_t$ stets automatisch eine Familie von Brown'schen Bewegungen $(B_\cdot^x)_{x\in\mathbb{R}}$ auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F})$ erhalten. Zudem, sind die induzierten Filtrierungen auf Ω unabhhängig vom Startpunkt, d.h. $\mathcal{F}_s^{B^x} = \mathcal{F}_s^{B^y} =: \mathcal{F}_s^B$ für alle $x, y \in \mathbb{R}$.

Somit können wir schließlich die *starke Markov-Eigenschaft* der Brown'schen Bewegung in der folgenden Form formulieren.

Satz 217. *Es sei $(B_t)_{t\geq 0}$ eine standard Brown'sche Bewegung definiert auf einem Wahrscheinlichkeitsraum Ω, und sei $\tau : \Omega \to \mathbb{R}$ eine endliche $\mathcal{F}_\cdot^B$-Stoppzeit sowie $F : \mathbb{R}^{\mathbb{R}_{\geq 0}} \to \mathbb{R}$ ein beschränktes messbares Pfadfunktional, dann gilt*

$$\mathbb{E}[F(B_{\tau+\cdot}^x)|\mathcal{F}_\tau^B] = \mathbb{E}_{B_\tau^x}[F(B_\cdot)] \textit{ fast sicher,}$$

wobei für $x \in \mathbb{R}$ der Prozess $(B_t^x)_{t\geq 0}$ definiert ist als $(x + B_t)_{t\geq 0}$.

Bemerkung 218. Die rechte Seite in der obigen Identität ist z.B. aufzufassen als die Zufallsvariable $H(B_\tau^x)$, welche durch Verknüpfung von B_τ^x mit der reellen Funktion $H : \mathbb{R} \to \mathbb{R}, z \to H(z) := \mathbb{E}[F(B_\cdot^z)]$ entsteht.

Beweis. Durch ein monotone Klassen-Argument führen wir die Aussage auf den Fall $F(B_.) = f(B_{t_1}, \cdots, B_{t_d})$ mit $f : \mathbb{R}^d \to \mathbb{R}$ stetig beschränkt zurück. Wir betrachten zunächst den Fall, dass der Wertebereich von τ endlich ist, d.h. das eine endliche Menge $S \subset \mathbb{R}_{\geq 0}$ existiert mit $\tau \in S$ fast sicher. Sei $Z : \Omega \to \mathbb{R}$ $\mathcal{F}^B_\tau$-messbar und beschränkt, dann gilt $Z1_{\tau=s} \in \mathcal{F}^B_s$, und daher mit der (elementaren) Markov-Eigenschaft

$$\begin{aligned}
\mathbb{E}[f(B_{\tau+t_1}, \cdots, B_{\tau+t_d})Z] &= \mathbb{E}\left[\sum_{s\in S} f(B_{s+t_1}, \cdots, B_{s+t_d})Z1_{\tau=s}\right] \\
&= \mathbb{E}\left[\sum_{s\in S} E_{B_s}[f(B_{t_1}, \cdots, B_{t_d})]Z1_{\tau=s}\right] \\
&= \mathbb{E}\left[E_{B_\tau}[f(B_{t_1}, \cdots, B_{t_d})]Z\right],
\end{aligned}$$

was zur Behauptung äquivalent ist. – Für den allgemeineneren Fall, dass τ beschränkt ist, approximieren wir durch die $\mathcal{F}^B_.$-Stoppzeiten $\tau^k := \frac{\lfloor k\tau+1 \rfloor}{k} \searrow \tau$ und erhalten zunächst mit dem obigen Argument, dass

$$\mathbb{E}[f(B_{\tau^k+t_1}, \cdots, B_{\tau^k+t_d})Z] = \mathbb{E}\left[E_{B^k_\tau}[f(B_{t_1}, \cdots, B_{t_d})]Z\right].$$

Für den Übergang zum Limes $k \to \infty$ Seiten verwenden wir schließlich noch auf der linken Seite dass $t \to B_t$ bzw. zusätzlich auf der rechten Seite dass für stetige Funktionen $f : \mathbb{R}^d \to \mathbb{R}$ die Funktion $z \to E_z F(f(B_{t_1}, \cdots, B_{t_d}))$ stetig sind. □

Bemerkung 219. Bei Beweis haben wir lediglich die Markov-Eigenschaft des Prozesses eingeschränkt auf endliche bzw. diskrete Zeitpunkte benutzt sowie im letzten Schritt die rechtsseitige Zeit-Stetigkeit der Trajektorien bzw. eine Stetigkeit des Erwartungswertes bezüglich dem Startpunkt. Beide Eigenschaften treffen für sehr viel mehr Markov-Prozesse als nur die Brown'sche Bewegung zu, so dass auch hier die starke Markov-Eigenschaft gilt.

Ganz analog zum obigen Vorgehen verläuft der Beweis der folgenden Aussage.

Satz 220. *Es sei B_t eine Brownsche Bewegung und τ eine fast sicher endliche $\mathcal{F}^B_.$-Stoppzeit. Dann ist $(B_{\tau+s} - B_s)_{s\geq 0}$ eine von F_τ unabhängige standard Brown'sche Bewegung.*

5.3.1. *Anwendung: Spieglungsprinzip.*

$$P\left(\sup_{0\leq s\leq t} B_s \geq a\right) = 2P\left(B_T \geq a\right)$$

und

$$\begin{aligned}
P\left(S_T \geq a\right) &= P\left(S_T \geq a, B_T \geq a\right) + P\left(S_T \geq, S_T < a\right) \\
&= P\left(B_T \geq a\right) + P\left(S_T \geq a, B_T < a\right)
\end{aligned}$$

$\tau := \inf_{s\geq 0}\{s \mid B_s \geq a\} \wedge 2T$

$$\begin{aligned}
P(S_T \geq a, B_T < a) &= P(\tau < T, B_T < a)\\
&= \mathbb{E}\left(1_{\tau<T}\cdot 1_{B_T<a}\right)\\
&= \mathbb{E}\left(\mathbb{E}\left(1\cdot \ldots \cdot 1\right) \mid \mathcal{J}_\tau\right)\\
&= \mathbb{E}_0\left(1_{\tau<T}\mathbb{E}_0\left(1_{B_T<a} \mid \mathcal{J}_\tau\right)\right)\\
&= \mathbb{E}_0\left(1_{\tau<T}\mathbb{E}_{B_\tau}\left(1_{B_{T-\tau}<a}\right)\right)\\
&= \mathbb{E}_0\left(1_{\tau(\omega)<T}\mathbb{E}_{X_{\tau(\omega)}(\omega)}\left(1_{B_{T-\tau(\omega)}(\omega')<a}\right)\right)\\
&= \mathbb{E}_0\left(1_{\tau(\omega)<T}\,\overbrace{\mathbb{E}_a\left(1_{B}\underbrace{T-\tau(\omega)}_{=s}(\omega')<a\right)}^{=\frac{1}{2}}\right)\\
&= \frac{1}{2}\mathbb{E}_0\left(1_{\tau(\omega)}\right)\\
&= \frac{1}{2}P\left(S_T > a\right)
\end{aligned}$$

6. Pfadeigenschaften der Brown'schen Bewegung

Erinnerung: $t \to f(t)$ ist lokal hölderstetig im Punkt t_0 mit Hölder-Index γ gdw.

$$\exists \delta > 0, c > 0, \forall t : |t - t_0| < \delta : |f(t) - f(t_0)| \leq C\,|t - t_0|^\gamma$$

- $\cdot$ $t \to f(t)$ lok. hölderstetig auf $I \subset \mathbb{R}$, gdw. $t \to f(t)$ lok. hölderstetig in allen $t_0 \in I$.
- $\cdot$ (Satz Teil1) $\forall \gamma \in [0, \frac{1}{2}) : t \to B_t(\omega)$ ist lok. h.-γ-st. auf $\mathbb{R}$ für P-fast jedes ω.
- $\cdot$ (Satz Teil2)$\forall \gamma > \frac{1}{2} : t \to B_t(\omega)$ nirgends lok. h.-st. auf $\mathbb{R}$ P-f.s..

Beweis. Teil1) Ersetze oBdA $I = \mathbb{R}_{\geq 0}$ durch $I = [0, 1]$. (Übergang zu $I = \mathbb{R}$ durch abzählbares Entfernen von Nullmengen) Lemma: $D := \left\{\underbrace{\frac{k}{2^n} \mid k = 0, ..., 2^n}_{=\bigcup_n D_n; D_n = \left\{\frac{k}{2^n}, k=0,...,2^n\right\}}, n \in \mathbb{N}\right\} \subset I$ Sei $f : D \to \mathbb{R}, m_0 \in \mathbb{N} : \forall m \geq m_0; \gamma > 0; \forall s, t \in D_m$ mit $|s - t| \leq 2^{-m}$:

$$|f(s) - f(t)| \leq 2^{-\gamma m}$$

Dann:

$$\forall s, t \in D, |s - t| \leq 2^{-m} : |f(s) - f(t)| \leq c|s - t|^\gamma$$

mit $c = 2^\gamma + \frac{2}{1-2^{-\gamma}}$. $\square$

Zusatz: Falls $f = \overline{f}_{|D}$ mit der Fortsetzung $\overline{f} : I \to \mathbb{R}$ und $\overline{f}$ stetig und $I := [0, 1]$. Dann ist $\overline{f} : I \to \mathbb{R}$ lok. h.-st. mit Exponent/Parameter γ.

Lemma 221. *$(B_s)_{s\geq 0}$ sei BB:*

$$\mathbb{E}\left(|X_{s+h}-X_s|^{2n}\right) \leq C_n h^2$$

$\alpha = 2n$, $\beta = n-1$, $\gamma = \frac{2n-1}{2n}$ mit bel. n.

Aussage 1) ist dann eine Konsequenz aus folgendem Satz:

Satz 222. *(von Kolm.-Chentsov) Sei $(X_t)_{t\geq 0}$ ein stetiger Prozess, d.h. $t \to X_t(\omega)$ st. für P-f.a. ω mit*

$$\forall s,t : \mathbb{E}\left(|X_s - X_t|^{\alpha}\right) \leq C|s-t|^{1+\beta}$$

mit $\alpha, \beta > 0$, so ist (X_s) lok. h.-st. f.s. für alle $\gamma \in [0, \frac{\beta}{2})$

Beweis. S_m sei die Menge aller benachbarten Punkte (s,t) in $D_m, \gamma < \frac{\beta}{\alpha}$.

$$A_m := \bigcup_{(s,t)\in S_m} \left\{|X_s - X_t| > 2^{-\gamma m}\right\}$$

$$P\left(A_m\right) \leq \sum_{(s,t)\in S_m} P\left(|X_s - X_t| > 2^{-\gamma m}\right)$$

Dann:

$$\begin{aligned} P\left(|X_s - X_t| > 2^{-\gamma m}\right) &= P\left(|X_s - X_t|^{\alpha} > 2^{-\alpha\gamma m}\right) \\ &\leq 2^{\alpha\gamma m}\mathbb{E}\left(|X_s - X_t|^{\alpha}\right) \\ &= 2^{\alpha\gamma m} C|s-t|^{-m(1+\beta)} \\ &= C2^{-m}2^{-\gamma m};\ \delta := \beta - \alpha\gamma > 0 \end{aligned}$$

es ex. ca. 2^m Paare (s,t) in D_m. Dann:

$$\begin{aligned} P\left(A_m\right) &\leq C2^m 2^{-\delta m} 2^{-m} \\ &= C2^{-\delta m} \end{aligned}$$

Dann:

$$\sum_m P\left(A_m\right) \leq C \sum 2^{-\delta m} < \infty$$

Nach Borel-Cantelli ist dann

$$P\left(\underline{\limsup} A_m\right) = 0$$

D.h. für P-f.a. ω gilt:

$$\omega \in A_m^c$$

für schließlich alle m, gdw. für P-f.a. ω ex. $m_0(\omega)$:

$$\omega \in A_m^c, \forall m \geq m_0$$

gdw.

$$\forall s,t \in D_m \text{ benachbart}, m \geq m_0 : |X_s(\omega) - X_t(\omega)| \leq 2^{-\gamma m}$$

Nach dem Lemma folgt dann, dass $s \to X_s(\omega)$ lok. h.-st. mit Parameter γ. □

Satz 223. *(Paley/Wiener/Zygmund)*

$$H_{\gamma,t} := \{f : [0,1] \to \mathbb{R} \mid f \text{ lok. h.-st. in } t \text{ mit Exponenten } \gamma\}$$

und

$$H_\gamma := \bigcup_{t\in[0,1]} H_{\gamma,t}$$

sei die Menge der Pfade, die irgendwo in $[0,1]$ lok. h.-st. sind. Zeige dann: $P(H_\gamma) = 0$.

Beweis. Falls $(\omega(t))_{t\geq 0} = B(t) =??? \in H_{\gamma,t}$, gdw. $\exists \delta > 0, C > 0 :$

$$|\omega_s - \omega_t| \leq C|s-t|^\gamma,\ \forall |s-t| \leq \delta$$

mit $\gamma > \frac{1}{2}$. Wähle $k \in \mathbb{N} : k > \frac{2}{2\gamma - 1}$; $n_0 = \lfloor \frac{k+1}{s} \rfloor$. Dann gilt für alle $n \geq n_0 :$ $i := \lfloor tn \rfloor + 1$, $\forall l \in \{0, ..., k-1\}$

$$\begin{aligned} \left|\frac{\omega_{i+l+1}}{n} - \frac{\omega_{i+l}}{n}\right| &= \left|\frac{\omega_{i+l+1}}{n} + \omega_t\right| + \left|\omega_t + \frac{\omega_{i+l}}{n}\right| \\ &\leq 2C\left(\frac{k+1}{n}\right)^\gamma \end{aligned}$$

Für $N \geq 2C(k+1)^\gamma, n \geq n_0(k) : i := i(n,t)$:

$$\omega \in H_{\gamma,t} \Longrightarrow \omega \in A_{N,n,i}$$

mit $A_{N,n,i} := \bigcap_{l=0}^{k-1} \left\{\omega \mid \left|\frac{\omega_{i+l+1}}{n} - \frac{\omega_{i+l}}{n}\right| \leq Nn^{-\gamma}\right\}$. Setze dann

$$\begin{aligned} A_{N,n} : &= \bigcup_{i=0}^{n} A_{n,N,i} \\ A_N^{(n_0)} : &= \bigcap_{n=n_0}^{\infty} A_{n,N} \\ A &= \bigcup_{N,n\in\mathbb{N}} A_N^{(n_0)} \end{aligned}$$

Das entspricht $H_\gamma \subset A$:

$$\mathbb{E}\left((B_t)^2\right) = t \iff B_t^2 = t \iff B = \sqrt{t}$$

$$\begin{aligned} P(B \in A_{N,n,i}) &= \left(P\left(|B_{1+n}| \leq Nn^{-\gamma}\right)\right)^k \\ &= \left(P\left(\left|\frac{1}{\sqrt{n}}B_1\right| \leq Nn^{-\gamma}\right)\right)^k \\ &= \left(P\left(|B_1| \leq Nn^{-\gamma+\frac{1}{2}}\right)\right)^k \\ &\leq 2^k N^k n^{k(\frac{1}{2}-\gamma)},\ \gamma > \frac{1}{2} \end{aligned}$$

$$\begin{aligned} P\left(B. \in A_N^{(n_0)}\right) &\leq \limsup_n P(A_{n,N}) \\ &\leq \lim_{n\to\infty} (n+1)2^k N^k n^{k(\frac{1}{2}-\gamma)} \\ &= 0 \end{aligned}$$

Dann ist $A_N^{(n_0)}$ eine Nullmenge. Dann ist A eine Vereinigung von Nullmengen und somit wieder Nullmenge. $\square$

7. Anhang: Hilfssaussagen aus der Masstheorie

7.1. **Monotone Klassen.** Wir reichen hier das sogenannte Monotone Klassen-Argument nach, das in vielen Beweisen benuzt wird, um Aussagen über σ-Algebren auf den Beweis von 'typischen' Elementen zu reduzieren. Analog kann man Aussagen über messbare Funktionen auf den Fall von 'typischen' messbaren Funktionen reduzieren. Am Ende des Abschnitts geben wir ein Beispiel anhand des Beweises vom Faktoriserungslemma.

Definition 224. Es sei Ω eine Menge und $\mathcal{M} \subset 2^\Omega$ ein Mengensystem auf Ω, so dass

(1) $\Omega \in \mathcal{M}$
(2) $A, B \in \mathcal{M}$ mit $B \subset A \Rightarrow A \setminus B \in \mathcal{M}$
(3) $A_1 \subset A_2 \subset \cdots$ mit $A_i \in \mathcal{M} \Rightarrow \cup_{i \in \mathbb{N}} A_i \in \mathcal{M}$,

dann heißt $\mathcal{M}$ ein *Dynkin-System.*[4] Ein Mengensystem $\mathcal{M}$ heißt *durchschnittsstabil*, falls

$$A, B \in \mathcal{M} \Rightarrow A \cap B \in \mathcal{M}.$$

Bemerkung 225.
(1) Jede σ-Algebra ist ein Dynkin-System; ein Dynkin-System ist eine σ-Algebra, sofern es durchschnittsstabil ist.
(2) Analog gilt wie für σ-Algebren, dass der Schnitt von Dynkin-Systemen wieder ein Dynkin-System ist. Entsprechend heißt für $\mathcal{M} \subset 2^\Omega$

$$\delta(\mathcal{M}) := \bigcap_{\substack{D\text{DynkinSystem} \\ \mathcal{M} \subset \mathcal{D}}} \mathcal{D}$$

das von $\mathcal{M}$ erzeugte Dynkin-System.

Satz 226 (Monotone Klassen-Argument)**.** *Sei $\Omega \neq \emptyset$ eine Menge.*

(1) *(Monotone Mengenklassen) Falls $\mathcal{M} \subset 2^\Omega$ ein durchschnittsstabiles Mengensystem ist, so gilt $\sigma(\mathcal{M}) = \delta(\mathcal{M})$, d.h. das von $\mathcal{M}$ erzeugte Dynkin-System ist gleich der erzeugten σ-Algebra.*
(2) *(Monotone Funktionenklassen) Falls $F \subset G$ zwei Vektorräume von reellwertigen Funktionen auf Ω sind, welche die konstanten Funktionen enthalten und die die beiden Voraussetzungen erfüllen, dass*
   (a) *$f \in F \Rightarrow |f| \in \mathcal{F}$ (bzw. alternativ $f \in F \Rightarrow f^2 \in F$)*
   (b) *$0 \leq g_n \nearrow g$ mit $g_n \in G$ und g beschränkt $\Rightarrow g \in G$,*
   so enthält G alls $\sigma(F)$-messbaren Funktionen.

Beweis. (1) Da jede σ-Algebra insbesondere ein Dynkin-System, gilt stets $\sigma(\mathcal{M}) \subset \delta(\mathcal{M})$. Wir müssen also nur noch zeigen, dass $\delta(\mathcal{M})$ bereits eine σ-Algebra ist. Hierzu zeigen wir, dass mit $\mathcal{M}$ auch das hiervon erzeugte $\delta(\mathcal{M})$ wieder durchschnittsstabil ist. Zu diesem Zweck sei $D \in \delta(\mathcal{M})$ fixiert und

$$\mathcal{D}_D := \{Q \in 2^\Omega \,|\, Q \cap D \in \delta(\mathcal{M})\},$$

[4]Diese Definition von Dynkin-System weicht von der äquivalenten Defintion aus W-Theorie I ab.

so überprüft man leicht, dass $\mathcal{D}_D$ wieder ein Dynkin-System ist. Aus der Durchschnitts-Stabilität von $\mathcal{M}$ folgt, dass $\mathcal{M} \subset \mathcal{D}_D$, also auch $\delta(\mathcal{M}) \subset \mathcal{D}_D$, d.h. also $D \cap D' \in \delta(\mathcal{M})$ für alle $D, D' \in \delta(\mathcal{M})$.
(ii) Wir führen hier den Beweis mit der ersten der beiden Voraussetzungen unter (a), d.h. mit $f \in F \Rightarrow |f| \in F$. Hieraus folgt, dass für $f, g \in F$ auch $f \wedge g = \frac{1}{2}(f+g-|f-g|)$ in F enthalten sind. Daher ist $\mathcal{M} := \{\{f \geq 1\} | f \in F\}$ ein durchschnittsstabiles Mengensystem auf Ω. Zudem gilt $\sigma(\mathcal{M}) = \sigma(F)$, denn $\{f \geq \alpha\} = \{f - \alpha + 1 \geq 1\}$ und $f - \alpha + 1 \in F$. Nach Voraussetzung (b) ist $\mathcal{D} := \{A \,|\, 1_A \in G\}$ ein Dynkin-System. Wir behaupten, dass $\sigma(\mathcal{M}) \subset \mathcal{D}$. Um dies zu zeigen, reicht es nach (1) zu zeigen, dass $\mathcal{M} \subset \mathcal{D}$. Sei hierfür $\{f \geq 1\} \in \mathcal{M}$ beliebig und $g := (f \wedge 1) \vee 0 \in F$, dann ist $0 \leq g \leq 1$ und $\{f \geq 1\} = \{g = 1\}$. Es gilt, dass $g^n \searrow 1_{\{g=1\}}$ auf Ω für $n \to \infty$. Die Funktion $t \to t^n$ kann als konvexe Funktion auf $\mathbb{R}$ dargestellt werden als punktweise gebildetes Supremum von abzählbaren vielen geeigneten affin-linearen Funktionen $(\phi_l)_{l \in \mathbb{N}}$, d.h. $t^n := \lim_{k \to \infty} \Phi_k(t)$ mit $\Phi_k(t) := \sup_{1 \leq l \leq k} \phi_l(t)$. Wegen der Stabilität von F unter dem (auf Ω) punktweise gebildeten Maximum gilt, dass $\Phi_k \circ g \in F$, und folglich wegen (b), dass $g^n = \sup_k \Phi_k \circ g \in G$. Erneut mit (b) folgt sodann aus $1 - g^n \nearrow 1 - 1_{\{g=1\}}$ für $n \to \infty$, dass auch $1_{\{g=1\}} \in G$, so dass also in der Tat $\{f \geq 1\} = \{g = 1\} \in \mathcal{D}$. Damit ist $\sigma(\mathcal{F}) \subset \mathcal{D}$ bewiesen. Hieraus und der Definition von $\mathcal{D}$ folgt nun, dass G alle elementaren $\sigma(F)$-messbaren Stufenfunktionen enthält, und erneut mit (b) durch monotone Approximation (ihres Positiv- und Negativteils) schließlich jede $\sigma(F)$-messbare Funktion. □

Korollar 227. *Es sei $(\Omega, \mathcal{F})$ ein messbarer Raum und $\mathcal{M} \subset 2^\Omega$ ein durchschnittsstabiles Mengensystem mit $\sigma(\mathcal{M}) = \mathcal{F}$. Sei V ein Vektorraum von reellen Funktionen auf Ω, so dass*

(1) *V enthält die konstanten Funktionen.*
(2) *Falls $0 \leq f_n \nearrow f$ mit $f_n \in V$ und f beschränkt $\Rightarrow f \in V$.*
(3) *Für jedes $A \in \mathcal{M}$ ist $1_A \in V$.*

Dann enthält V jede $\mathcal{F}$-messbare Funktion auf Ω.

Beweis. Folgt aus dem obigen Satz, angewendet auf F als linearer Hülle der Indikatorfunktionen zu den Mengen aus $\mathcal{M}$ und $G := V$. □

Als Beispiel für eine Anwendung bringen wir hier einen Spezialfall vom Faktoriserungslemma, das wir in der Vorlesung gelegentlich angesprochen haben.

Beweis vom Faktoriserungslemma (Satz 73). Wir führen die Aussage auf das obige Korollar zurück für den Fall $\mathcal{F} = \sigma(X) = \sigma(\mathcal{M})$ mit $\mathcal{M} = \{M = X^{-1}(A) | A \in \mathcal{E}\}$ und $V = \{Z = h \circ X, h : E \to \mathbb{R} \text{ messbar}\}$. Die Voraussetzungen (1) und (3) sind dann evidenterweise erfüllt. Für (2) sei $h_n(X) \to Z$ mit Z beschränkt so ist $h : E \to \mathbb{R}$, $h(e) := \limsup_{n \to \infty} h_n(e) \in [0, \infty]$ als numerische Funktion messbar, und es gilt $Z = h(X)$ wegen der Konvergenz von $h_n(e) \nearrow h(e) < \infty$ in allen Punkten $e \in X(\Omega) \subset E$. □

Bemerkung 228. Man kann auch einen elementaren Beweis für Satz 73 vermöge einer Approximation von Z durch Stufenfunktionen Z_n führen.